Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation
HAROLD BASOWITZ
New York University
EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University
JOHN M. BUTLER
University of Chicago
ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University
LOREN J. CHAPMAN
Southern Illinois University
JOHN H. CUMMING
New York State Department of Mental Hygiene
W. GRANT DAHLSTROM
University of North Carolina
JEROME D. FRANK
John Hopkins University
SOL L. GARFIELD
Columbia University
ANN MAGARET GARNER
Lincoln, Nebraska
ISRAEL GOLDIAMOND
Institute for Behavioral Research
HARRISON G. GOUGH
University of California, Berkeley
STARKE R. HATHAWAY
University of Minnesota
RALPH HEINE
University of Chicago
ERNEST R. HILGARD
Stanford University
LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University
SHELDON J. KORCHIN
University of California
WILLIAM G. LAWLOR
Fordham University
ARDIE LUBIN
U. S. Naval Hospital, San Diego
SALVATORE MADDI
University of Chicago
DAVID ROSENTHAL
National Institute of Mental Health
ALAN K. ROSENWALD
Illinois State Psychiatric Institute
SAUL ROSENZWEIG
Washington University, St. Louis
JACQUES RUTSCHMANN
New York State Psychiatric Institute
SEYMOUR SARASON
Yale University
JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University
ROY SCHAFER
Yale University
JOHN SHLIEN
University of Chicago
MORRIS I. STEIN
New York University
CHARLES STROTHER
University of Washington
HAROLD L. WILLIAMS
University of Oklahoma School of Medicine
C. L. WINDER
Michigan State University
JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

FRANCES L. BREWER
Editorial Assistant

VOLUME 70
1965

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Prince and Lemon Sts., Lancaster Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Copyright 1965 by the American Psychological Association, Inc.

ACKNOWLEDGMENT

The Editorial Staff wishes to express its thanks and appreciation to the following individuals who aided in the reviewing process as *ad hoc* consultants during 1964–65:

Arnold Mordkoff
Benson Ginsburg
Reginald Taylor
John Gibbon
W. Crawford Clark
Rue Cromwell
Kurt Salzinger
Howard Roffwarg
Joseph Jaffe
James Barnard

Bureau Edni. [illegible] Research
[illegible] COLLEGE
Dated 8. 9. 70
Accs. No. J — 422

CONTENTS OF VOLUME 70

ARTICLES

PAGE

Amarel, Marianne, & Cheek, Frances E. **Some Effects of LSD-25 on Verbal Communication** 453
Arenberg, David. **Anticipation Interval and Age Differences in Verbal Learning** 419
Barber, Theodore Xenophon. **Experimental Analysis of "Hypnotic" Behavior: A Review of Recent Empirical Findings** 132
Bishop, Patricia F. See Lorr, Maurice.
Blacker, K. H. See Hargreaves, William A.
Blinder, Barton. See Luborsky, Lester.
Blumenthal, Richard, Meltzoff, Julian, & Rosenberg, Selig. **Some Determinants of Persistence in Chronic Schizophrenic Subjects** 246
Bryan, James H. See Lichtenstein, Edward.
Buss, Arnold H., & Lang, Peter J. **Psychological Deficit in Schizophrenia: I. Affect, Reinforcement and Concept Attainment** 2
Buss, Arnold H. See Lang, Peter J.
Butterfield, Earl C., & Zigler, Edward. **The Effects of Success and Failure on the Discrimination Learning of Normal and Retarded Children** 25
Byrne, Don, & Sheffield, John. **Response to Sexually Arousing Stimuli as a Function of Repressing and Sensitizing Defenses** 114
Cheek, Frances E. See Amarel, Marianne.
Cromwell, Rue L. See Dokecki, Paul R.
Davidson, Gerald C. See Lazarus, Arnold A.
Di Loreto, Adolph. See Saltz, Eli.
Dohrenwend, Barbara Snell. See Dohrenwend, Bruce P.
Dohrenwend, Bruce P., & Dohrenwend, Barbara Snell. **The Problem of Validity in Field Studies of Psychological Disorder** 52
Dokecki, Paul R., Polidoro, Lewis G., & Cromwell, Rue L. **Commonality and Stability of Word Association Responses in Good and Poor Premorbid Schizophrenics** 312
Donovan, M. J., & Webb, Warren W. **Meaning Dimensions and Male-Female Voice Perception in Schizophrenics with Good and Poor Premorbid Adjustment** 426
Edmonston, William E., Jr. See Plapp, Jon M.
Elkind, David, Koegler, Ronald R., Go, Elsie, & Van Doorninck, William. **Effects of Perceptual Training on Unmatched Samples of Brain-Injured and Familial Retarded Children** 107
Ellsworth, Robert B. **A Behavioral Study of Staff Attitudes toward Mental Illness** .. 194
Eriksen, Charles W. See Williamsen, John A.
Farina, Amerigo, & Ring, Kenneth. **The Influence of Perceived Mental Illness on Interpersonal Relations** 47
Fiske, Donald W., & Goodman, Gerald. **The Posttherapy Period** 169
Flenning, Frank. See Rosenbaum, Gerald.
Folsom, Angela T. See Hamlin, Roy M.
Foulkes, David, & Vogel, Gerald. **Mental Activity at Sleep Onset** 231
Fox, Robert. **Rate of Binocular Rivalry Alternation in Psychotic and Nonpsychotic Patients** 34
Go, Elsie. See Elkind, David.
Goldman, Arnold R. **Differential Effects of Social Reward and Punishment of Dependent and Dependency-Anxious Schizophrenics** 412
Goldstein, Robert H., & Salzman, Leonard F. **Proverb Word Counts as a Measure of Overinclusiveness in Delusional Schizophrenics** 244
Goodenough, Donald R. See Karp, Stephen A.
Goodman, Gerald. See Fiske, Donald W.
Goodstein, Leonard D. See Spence, Janet Taylor.
Gough, Harrison G. **Conceptual Analysis of Psychological Test Scores and Other Diagnostic Variables** 294
Gough, Harrison G., Wenk, Ernest A., & Rozynko, Vitali V. **Parole Outcome as Predicted from the CPI, the MMPI, and a Base Expectancy Table** 432
Haber, Merry Marquit. See Zuckerman, Marvin.
Hamlin, Roy M., Haywood, H. Carl, & Folsom, Angela T. **Effect of Enriched Input on Schizophrenic Abstraction** 390
Hare, Robert D. **Temporal Gradient of Fear Arousal in Psychopaths** 442
Hargreaves, William A., Starkweather, J. A., Blacker, K. H. **Voice Quality in Depression** 218
Haywood, H. Carl. See Hamlin, Roy M.
Hess, Eckhard H., Seltzer, Allan L., & Shlien, John M. **Pupil Response of Hetero- and Homosexual Males to Pictures of Men and Women: A Pilot Study** 165

Higgins, Jerry, Mednick, Sarnoff A., & Philip, Franklin J. Associative Disturbance as a Function of Chronicity in Schizophrenia 451
Honigfeld, Gilbert. Temporal Effects of LSD-25 and Epinephrine on Verbal Behavior 303
Hunt, Howard F., & Thetford, William N. Editorial 1
Hunt, William A., Schwartz, Melvin L., & Walker, Ronald E. Reliability of Clinical Judgments as a Function of Range of Pathology 32
Huse, Mary M., & Parsons, Oscar A. Pursuit-Rotor Performance in the Brain Damaged 350
Jellinek, E. Morton. See Shakow, David.
Jenkins, James J., & Schuell, Hildred. Acquiescence Response Set in Aphasics 111
Johnson, Harold J. See Williamsen, John A.
Johnston, Roy, & McNeal, Benjamin F. Residual Psychopathology in Released Psychiatric Patients and Its Relation to Readmission 337
Karp, Stephen A., Witkin, Herman A., & Goodenough, Donald R. Alcoholism and Psychological Differentiation: Effect of Alcohol on Field Dependence 262
Kates, Frances F. See Kates, Solis L.
Kates, Solis L., & Kates, Frances F. Social and Nonsocial Verbal Concepts of Deaf and Hearing Children 214
Kennedy, Wallace A. School Phobia: Rapid Treatment of Fifty Cases 285
Koegler, Ronald R. See Elkind, David.
Lair, Charles V. See Spence, Janet Taylor.
Lair, Charles V. See Spence, Janet Taylor.
Lang, Peter J., & Buss, Arnold H. Psychological Deficit in Schizophrenia: II. Interference and Activation 77
Lang, Peter J., Lazovik, A. David, & Reynolds, David J. Desensitization, Suggestibility and Pseudotherapy 395
Lang, Peter J. See Buss, Arnold H.
Lazarus, Arnold A., Davison, Gerald C., & Polefka, David A. Classical and Operant Factors in the Treatment of a School Phobia 225
Lazovik, A. David. See Lang, Peter J.
Lee-Teng, Evelyn. Trance-Susceptibility, Induction-Susceptibility, and Acquiescence as Factors in Hypnotic Performance 383
Leonard, James R. Hypnotic Age Regression: A Test of the Functional Ablation Hypothesis 266
Lichtenstein, Edward, & Bryan, James H. Acquiescence and the MMPI: An Item Reversal Approach 290
London, Perry. See Slotnick, Robert.
Lorr, Maurice, Bishop, Patricia F., & McNair, Douglas M. Interpersonal Types among Psychiatric Patients 468
Luborsky, Lester, Blinder, Barton, & Schimek, Jean. Looking, Recalling, and GSR as a Function of Defense 270
Manasse, Gustave. Self-Regard as a Function of Environmental Demands in Chronic Schizophrenics 210
McNair, Douglas M. See Lorr, Maurice.
McNeal, Benjamin F. See Johnston, Roy.
Mednick, Sarnoff A. See Higgins, Jerry.
Megargee, Edwin I. Relation between Barrier Scores and Aggressive Behavior 307
Meltzoff, Julian. See Blumenthal, Richard.
Merbaum, Michael, & Southwell, Eugene A. Conditioning of Affective Self-References as a Function of the Discriminative Characteristics of Experimenter Intervention 180
Parloff, Morris B. See Wispé, Lauren G.
Parsons, Oscar A. See Huse, Mary M.
Philip, Franklin J. See Higgins, Jerry.
Plapp, Jon M., & Edmonston, William E., Jr. Extinction of a Conditioned Motor Response Following Hypnosis 378
Polefka, David A. See Lazarus, Arnold A.
Polidoro, Lewis G. See Dokecki, Paul R.
Quarrington, Bruce. Stuttering as a Function of the Information Value and Sentence Position of Words 221
Reynolds, David J. See Lang, Peter J.
Riedel, Wolfgang W. Anxiety Level and the "Doubtful" Judgment in a Psychophysical Experiment 462
Ring, Kenneth. See Farina, Amerigo.
Rosen, Harold. See Rosenbaum, Gerald.
Rosenbaum, Gerald, Flenning, Frank, & Rosen, Harold. Effects of Weight Intensity on Discrimination Thresholds of Normals and Schizophrenics 446
Rosenberg, Selig. See Blumenthal, Richard.
Rozynko, Vitali V. See Gough, Harrison G.
Saltz, Eli, & Di Loreto, Adolph. "Defense" against Traumatic Concepts 281

Salzman, Leonard F. See Goldstein, Robert H.

Sapolsky, Allan. **Relationship between Patient-Doctor Compatibility, Mutual Perception, and Outcome of Treatment** 70

Schimek, Jean. See Luborsky, Lester.

Schuell, Hildred. See Jenkins, James J.

Schwartz, Melvin L. See Hunt, William A.

Seltzer, Allan L. See Hess, Eckhard H.

Shakow, David, & Jellinek, E. Morton. **Composite Index of the Kent-Rosanoff Free Association Test** 403

Sheffield, John. See Byrne, Don.

Sherman, James A. **Use of Reinforcement and Imitation to Reinstate Verbal Behavior in Mute Psychotics** 155

Shlien, John M. See Hess, Eckhard H.

Slotnick, Robert, & London, Perry. **Influence of Instructions of Hypnotic and Nonhypnotic Performance** 38

Smith, Colin M. **Family Size in Alcoholism** 230

Southwell, Eugene A. See Merbaum, Michael.

Spence, Janet Taylor, Goodstein, Leonard D., & Lair, Charles V. **Rote Learning in Schizophrenic and Normal Subjects under Positive and Negative Reinforcement Conditions** 251

Spence, Janet Taylor, & Lair, Charles V. **Associative Interference in the Paired-Associate Learning of Remitted and Noremitted Schizophrenics** 119

Starkweather, J. A. See Hargreaves, William A.

Stoyva, Johann Martin. **Finger Electromyographic Activity During Sleep: Its Relation to Dreaming in Deaf and Normal Subjects** 343

Tecce, Joseph J. **Relationship of Anxiety (Drive) and Response Competition in Problem Solving** 465

Thetford, William N. See Hunt, Howard F.

Van Doorninck, William. See Elkind, David.

Vogel, Gerald. See Foulkes, David.

Walker, Ronald E. See Hunt, William A.

Webb, Warren W. See Donovan, M. J.

Weinstein, Laura. **Social Schemata of Emotionally Disturbed Boys** 457

Wenk, Ernest A. See Gough, Harrison G.

Williamsen, John A., Johnson, Harold J., & Eriksen, Charles W. **Some Characteristics of Posthypnotic Amnesia** 123

Willner, Allen. **Impairment of Knowledge of Unusual Meanings of Familiar Words in Brain Damage and Schizophrenia** 405

Wispé, Lauren G., & Paroff, Morris B. **Impact of Psychotherapy on the Productivity of Psychologists** 188

Witkin, Herman A. **Psychological Differentiation and Forms of Pathology** 317

Witkin, Herman A. See Karp, Stephen A.

Wolowitz, Howard M. **Attraction and Aversion to Power: A Psychoanalytic Conflict Theory of Homosexuality in Male Paranoids** 360

Worell, Leonard. **The Ring of Punishment: A Theoretical and Experimental Analogue of Repression-Suppression** 201

Volume 70 February 1965 Number 1

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE BURDOCK
New York State Psychiatric Institute

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

NICHOLAS HOBBS
George Peabody College

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

FRANCES L. BREWER
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 1

EDITORIAL

The *Journal of Abnormal Psychology* will give priority to articles on problems related to abnormal behavior, broadly defined. The journal's interests thus include the following: (*a*) psychopathology—its development or acquisition, its treatment or remission, and its symptomatology and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional or behavioral pathology; (*e*) social or group effects on adjustment or pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. This list does not pretend to be exhaustive; nor does the order of listing imply an order of priority. The boundaries of the journal's interests cannot be hard and fast and, in fact, probably will be breached on occasion at the discretion of the editors and consultants.

As a matter of general policy, all articles accepted must add to knowledge in some way, whether they are experimental reports, case studies, or theoretical papers.

Experimental studies should be of sufficiently high quality to make a specific contribution—to discover something. A paper based primarily on simple enumeration of the characteristics of subjects in some diagnostic category, but which does not present hypotheses and comparisons with other groups in order to substantiate hypotheses, is not acceptable. Studies with negative results are acceptable provided they disconfirm an expectation or hypothesis of some interest, and one that had a reasonable antecedent probability of being true. Such studies must be sound enough to provide a decisive disconfirmation, of course, rather than simply "fail to confirm" expectations (as, say, with tests of low reliability).

Long experimental articles should present a substantial body of data, generally such as would be derived from a series of interlocking experiments, rather than report a single short experiment in an overly-expanded way. *Short experimental notes* must be solidly based experiments, but need not be of the same scope as the longer articles.

Reviews and theoretical papers will be accepted provided they present new ideas, offer a revision of existing views, or organize ideas, phenomena, or problems in new and, hopefully, fertile ways. Because such papers tend to be lengthy, they can be accepted only if they are on topics of particular pertinence to abnormal psychology. *Short, critical notes* in the scholarly tradition may be accepted, on occasion, provided they bring new thoughts and views to bear on a problem or experimental finding. Notes that are essentially "letters to the editor," however, will not be considered.

Case studies have a traditional place in the journal and will be accepted when they are of high quality. The problem here, as with reviews, is that such articles tend to be long. Because of this, the requirements for acceptance are similar—exposition of good quality, new organizations and juxtapositions of ideas and phenomena, and potential for generating fertile approaches to problems in this area. If the number of acceptable case studies becomes too great to handle within the regular format, the possibility of publishing case study supplements from time to time will be explored.

It is hoped and expected that every issue will contain a mixture of these different kinds of articles. Short experimental reports of restricted scope, following the habits and traditions of contemporary psychology, are always welcome. To achieve balance and add depth to the journal, however, we also want to extend a most cordial invitation to longer, more substantial papers that grapple with larger issues in a more decisive way, with the caveat that while a paper does not have to be long to be good, it does have to be good to be long.

Howard F. Hunt
William N. Thetford

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 2–24

PSYCHOLOGICAL DEFICIT IN SCHIZOPHRENIA:

I. AFFECT, REINFORCEMENT, AND CONCEPT ATTAINMENT

ARNOLD H. BUSS AND PETER J. LANG[1]

University of Pittsburgh

This, the 1st of 2 papers reviewing laboratory studies of psychological deficit in schizophrenia, considers the effects of social censure, affective stimuli, cooperation, urging, and verbal and nonverbal reinforcers on performance on a variety of tasks. Thus, disturbance in concept formation is evaluated in terms of: loss of abstract attitude, communication disturbance, regression, and the consequence of erratic attention. Finally, various theories of schizophrenic deficit are evaluated. Although motivational constructs cannot be wholly dismissed, interference theory is generally more comprehensive and parsimonious. Many specific hypotheses of interference theory still need to be substantiated by experiment, however.

The term *psychological deficit* was coined by Hunt and Cofer (1944) who wanted a neutral phrase to describe the decrement shown by psychiatric patients in comparison to normals on various laboratory and intellectual tasks. They reviewed the substantial body of research accumulated prior to World War II.

Since that time there has been a prolific output of research on deficit, too much to encompass within a single review. Therefore we shall consider only laboratory studies of psychological deficit in schizophrenia, omitting reports of tests and clinical observations.

Organizing the voluminous literature proved to be a difficult task. One possibility was to group studies on the basis of the major theoretical approaches to deficit, since theories of deficit are essentially theories of schizophrenia; but this would have led to excessive repetition. Therefore the literature has been organized around seven areas defined in part by theory but more broadly in terms of issues and methods of investigation: (*a*) affect and reinforcement, (*b*) concept attainment, (*c*) attention, (*d*) set, (*e*) associative interference, (*f*) drive, and (*g*) somatic arousal.

The dominant orientation in the affect and reinforcement area might be called a social-motivational view, which has two main variants. One suggests that schizophrenics are oversensitive to punishing and/or affective stimuli, which cause them to withdraw and which produce performance deficit. The other assumes that schizophrenics are already so withdrawn from interpersonal situations that the usual incentives, rewards, and punishments employed in experimental situations do not motivate them, and their performance suffers.

In the concept attainment area, research relevant to four explanatory concepts is reviewed. The first approach holds that both schizophrenics and brain-damaged patients suffer from a loss of the "abstract attitude." A second view argues that deficit is attributable to a communication disturbance, and a third emphasizes regression: later, more mature modes of functioning are ostensibly given up for earlier, more concrete, and less efficient modes. Finally, more recent research suggests that differences in the concepts of normals and schizophrenics occur because the latter are distractible and respond to irrelevant cues.

Research on attention, set, and association further emphasizes the importance of interference effects in deficit. The schizophrenic is regarded as being excessively distracted by both incidental, external stimuli and intrusive associations, as failing to maintain a proper set or orientation, and as failing to alter the set when such a change is appropriate.

Two drive theories are evaluated. The better known theory assumes that the schizophrenic is extremely anxious, this anxiety being a high drive state that worsens performance, especially on complex tasks. The lesser known theory suggests that in schizophrenia there is a protective (cortical) inhibition that slows down learning and increases reminiscence. Interpretations of deficit in terms of the neurophysiological concept of arousal are also considered, and the relevant research on the somatic responses of schizophrenics is evaluated.

---

[1] Now at the University of Wisconsin.

The exposition is divided into two papers. In this first paper we review the affect-reinforcement and concept-regression areas. A subsequent paper considers research on interference effects (attention, set, and associative interference) and activation (drive and arousal) as well as general problems of method in deficit research.

AFFECT AND REINFORCEMENT

Social Censure

Rodnick and Garmezy (1957) have been the most forceful advocates of the censure hypothesis:

> We have indicated that clinical psychiatric reports stress the sensitivity of the schizophrenic patient to the threat of criticism or rebuff inherent in almost any social situation. If such criticism does accentuate the patient's difficulty in differentiating cues in his environment, then it would follow that the experimental introduction of censure should produce greater discrimination decrements in schizophrenic patients than in normal individuals [p. 118].

Rodnick and Garmezy have attempted to confirm the censure hypothesis in a series of experiments conducted by themselves or by their students. Garmezy (1952) showed that after censure (the word WRONG flashed after a response) schizophrenics had flatter generalization gradients and more errors than after praise (the word RIGHT flashed after a response); the performance of normals improved with censure. Webb (1955) found that failure led to poorer performance on a conceptual task, whereas a control group *improved* in their performance. All subjects were schizophrenics, and it is not known whether failure would have led to poorer performance in normals.

Bleke (1955) compared normals, good premorbid, and poor premorbid schizophrenics on a memory task. The subjects were either rewarded (RIGHT) or punished (WRONG) during learning. The poor premorbid schizophrenics manifested significantly better reminiscence and relearning than did good premorbids or normals. This was interpreted in terms of the effect of censure: criticism interfered with the learning of the poor premorbids, but on later testing for recall in the absence of censure they improved.

Smock and Vancini (1962) followed up Bleke's study, using a similar task but without differentiating between good and poor premorbid schizophrenics. After a practice session, half the subjects were rewarded (told they had done well), and half were censured (told they had done poorly). The practice was followed by learning without reinforcement, and then recall. There were no differences between normals and schizophrenics or between praised and censured subjects in original learning, but the censured schizophrenics had significantly *less* reminiscence than both the praised schizophrenics and the normal subjects. These findings are opposed to Bleke's, but there are differences between the studies that might be important: the makeup of the schizophrenic sample (good and poor premorbid versus a mixed group), the task, and the way praise and censure were delivered (after each response versus after practice trials). The discrepancy in findings and the possible reasons for it suggest that the effects of praise and censure may be too complex to be handled within the Rodnick-Garmezy framework.

Alvarez (1957) had normals, good premorbids, and poor premorbids make judgments about pictures. The judgments were either censured (WRONG) or praised (RIGHT), and later preference for the pictures was assessed. Poor premorbid schizophrenics were more susceptible to the effects of censure than good premorbids or normals, that is, the poor premorbids showed the greatest decline in preference for the pictures associated with WRONG. Thus the poor premorbids were more sensitive to WRONG, as judged by their preference scores; note that this does not mean that WRONG disrupted their performance, for this was not assessed. Similar results have been reported by Neiditch (1963), who found that functional psychotics lowered their preference for tasks which followed a failure situation significantly more than normals. He interpreted these data in terms of an inappropriate set.

Finally, Zahn (1959) tested the perceived size of pictures denoting scolding and feeding of a child by a mother. The size estimates were rewarded (RIGHT) for some subjects and punished (WRONG) for others. Good premorbid schizophrenics tended to *overestimate* the size of the punished pictures, whereas normals and poor premorbids did not.

> The results are interpreted in terms of a high degree of anxiety or affective responsivity in the Goods and the predominance of avoidance and withdrawal mechanisms in the Poors [Zahn, 1959, pp. 46–47].

Most of these experiments were conducted under the aegis of Rodnick and Garmezy (1957). Taken at face value, they offer some support for the social censure hypothesis: failure or being told WRONG seems to produce schizophrenic deficit, an effect more striking in poor premorbid than in good premorbid schizophrenics. However, the evidence is weak, and individual experiments are open to a variety of interpretations. We shall return to these issues in the larger context of reward and punishment.

"Affective Stimuli"

Censure stimuli. The social censure hypothesis has a corollary pertaining to stimuli: schizophrenics are oversensitive to, and therefore disrupted by, stimuli connoting parental punishment or negative parent figures. This corollary has been tested mainly by students of Rodnick and Garmezy.

Dunn (1954) presented pictures of scolding, whipping, feeding, and neutral objects to schizophrenics and normals. For each scene there was a standard picture and five variations; the task was to judge whether the standard and the variations (presented in pairs) were the same or different. The schizophrenics judged more of the scolding variations as being the same as the standard than did the normals, but the other scenes yielded no differences. Dunn also discovered that a history of maternal conflict led to more judgments of "same" on the scolding scene. Two issues make it difficult to interpret these findings. First, the scolding scene is less clearly represented and might depict a number of different interactions, whereas the raised whip and the food in hand leave no doubt about the relationships presented. Thus ambiguity might account for the results, assuming that schizophrenics have more difficulty with complex stimuli which elicit many alternate responses. Second, if it were established that scolding (but not whipping or feeding) is the major trauma in the parent-child interactions of schizophrenics, the social censure hypothesis might be supported by these data. However, there is at present no basis for assuming that scolding is a more prominent problem for schizophrenics than whipping or feeding difficulties.

Dunn's results are only partially corroborated by those of Turbiner (1961), whose procedure was almost identical. In addition to neutral pictures, Turbiner had pictures of scolding and of affection. The schizophrenics' judgments were the same as the normals' for neutral stimuli but poorer than normals' for *both* positive and negative affective pictures.

Garmezy and Rodnick (1959) suggested that good and poor premorbid schizophrenics are not equally sensitive to censure from both parents. Specifically, they proposed that good premorbid schizophrenics are disrupted by paternal censure and poor premorbids by maternal censure. Two of their students, Dunham (1959) and Kreinik (1959), attempted to verify this hypothesis.

Dunham repeated Dunn's experiment, with minor modifications. Compared to normal subjects, good premorbid schizophrenics showed deficit only on slides depicting father censure and poor premorbids showed deficit on both mother censure and neutral slides. These are positive findings, but the poor premorbids' deficit with neutral slides is difficult to reconcile with the hypothesis.

Kreinik's experiment is open to two interpretations. She employed a conceptual task in which all subjects were presented with stimuli connoting positive fathers, negative fathers, positive mothers, negative mothers, and neutrality (nonhuman). There were two sessions, and only two groups failed to show improvement in the second session. These were the good premorbids with positive father stimuli and the poor premorbids with positive mother stimuli. Kreinik attributed these results to sequence effects. The Goods presented with negative father stimuli in the first session failed to improve with positive father stimuli in the second session. The results were the same for Poors with mother stimuli. Kreinik assumed that the deleterious effects of the negative parental stimuli carried over to the session with the positive stimuli. The problem with this interpretation is that the sessions were 1 day apart, and it appears doubtful that any effects would be sustained for 24 hours. The alternate explanation is more parsimonious: Goods and Poors are negatively affected by positive father and mother figures, respectively.

Two other studies are relevant. Baxter and Becker (1962) used TAT cards depicting mother-son and father-son relationships, scoring the stories for anxiety. Poor premorbids scored higher in anxiety than good premorbids on the mother-son picture, and the reverse was true for the father-son picture. However, two issues make it difficult to interpret these results unequivocally. First, the Goods scored approximately the same in anxiety on the mother-son and father-son pictures. Second, there was no control group of normal subjects, which means we do not know how much anxiety ordinarily appears in stories about these pictures.

Lebow and Epstein (1963) used TATlike stimuli depicting mother, father, and peer interactions in nurturant, rejecting, and ambiguous situations. The subjects were good premorbid schizophrenics and medical patients. In the present context the only relevant measure was a rating of the stories, called goodness of response. Concerning this measure,

> there was a strong tendency for the schizophrenics to tell their poorest stories to the nurturant cues and their best stories to the ambiguous cues. Thus, if anything, the schizophrenics find nurturant as well as rejecting cues disruptive, and the former more so. Rather than a specific censure-cue deficit, the schizophrenic exhibits a general deficit for cues associated with emotional involvement, whether of a positive or a negative nature [p. 32].

Size estimation has also been used as an index of the distorting effects of affective stimuli. Harris (1957) presented pictures of the mother-son interaction. Normals made few mistakes in size estimation, and good premorbids underestimated the size of the pictures. In a similar study Zahn (1959) tested perceived size of scolding and feeding pictures. In addition the pictures were associated with RIGHT for some subjects and with WRONG for others. Good premorbids, in relation to the rewarded and the feeding pictures, overestimated the size of scolding pictures and pictures associated with WRONG. There were no similar relationships for normals or poor premorbids. In attempting to explain the discrepancy (Harris' good premorbids *under*estimated, Zahn's good premorbids *over*estimated), Zahn pointed out that his schizophrenics had been hospitalized longer than Harris', although it is not clear how this accounts for the discrepancy.

Moriarty and Kates (1962) matched normals, good premorbid, and poor premorbid schizophrenics for concept attainment with neutral stimuli and tested them with stimuli connoting parental approval and disapproval. For *both* approval and disapproval stimuli the normals performed significantly better than the good premorbids, who performed significantly better than the poor premorbids. Evidently, "social" stimuli disrupted the performance of the schizophrenics. Concerning specific effects of parental disapproval,

> all groups show deficit behavior on social materials suggesting censure, with the psychiatric and the normal groups showing about the same relative decrements on the disapproval cards [p. 362].

Taken together, these studies do not support the hypothesis that the performance of schizophrenics is disrupted by stimuli connoting censure, especially parental censure. Positive findings have been reported, but they have been open to alternate interpretations because of methodological problems (e.g., Dunn, 1954; Kreinik, 1959) or unclear results (e.g., Dunham, 1959; Baxter & Becker, 1962). Early positive results have been either contradicted by later results (e.g., Harris, 1957; Zahn, 1959) or only weakly corroborated by later results (e.g., Dunn, 1954; Turbiner, 1961). There have also been negative findings: schizophrenics are adversely affected not only by censure stimuli but also by stimuli connoting affection, nurturance, and approval.

Affective stimuli in general. If the social censure hypothesis is correct, deficit should be produced *only* by stimuli associated with censure or parental punishment. An alternate hypothesis is that schizophrenics are oversensitive to *all* affective stimuli, not merely those connoting censure. A number of studies have attempted to test this hypothesis.

Culver (1961) had subjects estimate the size of pictures of a mother and a neutral object. *Both* good and poor premorbids underestimated the size of the mother, whereas normals did not. All three groups were similar in size estimation of the square. Raush (1956) tested for perceived size of a neutral object (overcoat button) and symbolic objects (ice cream cone and cigar). Using the estimated size of the neutral object as a base, he found that schizophrenics overestimated the size of the symbolic objects in comparison to the normals. Ehrenworth (1960) had subjects estimate the relative size of drawings of geometric figures and themes of heterosexuality, autonomy, competition, affiliation, and authority (all affective themes). Schizophrenics' size estimates of the affective stimuli were significantly poorer than both their estimates of the neutral stimuli and normals' estimates of the affective stimuli.

Pishkin, Smith, and Leibowitz (1962) used figures dressed as psychiatric aides or patients, and the subjects judged the relative size of the figures versus a silhouette. For the aide figure the errors of schizophrenics and normals were similar; for the patient figure the schizophrenics made significantly more errors. Assuming the patient figure was more symbolic, Pishkin et al. concluded that

> it is the symbolic or emotional value of the stimulus which accounts for schizophrenic deficit on perceptual tasks [p. 329].

De Wolfe (1962) showed that good premorbids are more emotionally responsive to affective words than both poor premorbids and normals in a verbal conditioning task. Feldstein (1962) tested speech disturbance to affective and neutral materials and found no differences between schizophrenics and normals.

Four studies reported significant results using the human-nonhuman dichotomy. Davis and Harrington (1957) matched subjects on the basis of their ability to utilize information about nonhuman stimuli in a conceptual-discrimination problem. Schizophrenics were severely disrupted by human stimuli in a comparable problem, whereas normals were not. Marx (1962) divided his schizophrenic subjects into "early" and "late" patients (similar to acute versus chronic). In a conceptual task, human stimuli disrupted the early schizophrenics more than the late schizophrenics. Whiteman (1954) had subjects learn formal concepts (based on physical properties or size relationships) and social concepts (based on interactions among people, such as cooperation or encouragement). Although the schizophrenics were poorer than normals on the formal, nonhuman concepts, their performance significantly worsened on the social, human concepts. Brodsky (1961) compared schizophrenics and nonpsychotic hospital patients on a conceptual task involving people and nonhuman stimuli. The schizophrenics' conceptual performance was less adequate than normals' only with human stimuli.

Feffer (1961) tested the perception of affective and neutral words. He divided his schizophrenic subjects into "pathologically-concrete" and "adequately-conceptualizing" groups on the basis of tests of conceptual thinking. On the perceptual task the adequately-conceptualizing schizophrenics were especially vigilant in the presence of affective stimuli; the pathologically-concrete schizophrenics tended to avoid affective stimuli; and the normals showed no particular reaction to affective stimuli.

Silverman (1963) reported similar findings when words were presented tachistoscopically. Schizophrenics were especially sensitive to unpleasant words, identifying them more correctly than did normals. There were two kinds of schizophrenics, paranoid and nonparanoid. The paranoids placed between nonparanoids and normals in their differential sensitivity to unpleasant words. These results raise the possibility that Feffer's pathologically-concrete and adequately-conceptualizing schizophrenics may have been nonparanoid and paranoid, respectively.

One study failed to find a difference in perception between schizophrenics and normals. Nelson and Caldwell (1962) measured depth perception, using as stimuli drawings of a man, a woman, a dog, and a circle. There were no significant differences in perception for different stimuli or between normals and schizophrenics.

Arey (1960) tested subjects for recognition of sexual and neutral pictures. He found that schizophrenics were more affected by the sexual pictures, i.e., their recognition performance on sexual pictures compared to nonsexual pictures was significantly poorer than normals.

These various studies indicate that schizophrenic deficit occurs in response to many different kinds of stimuli, not merely those connoting censure. Thus the censure hypothesis is not sufficiently encompassing to account for the data. The results can be better explained by the alternate hypothesis that schizophrenics are oversensitive to all affective stimuli.

This latter hypothesis suggests that schizophrenics are disturbed by certain kinds of stimuli and that such disturbance leads to a deterioration in performance. The central problem is to define the term *affective,* whose meaning appears to change from one experiment to the next. It can have taboo aspects, as in the sexual stimuli used by Arey (1960); it may refer to human, as opposed to nonhuman stimuli; it may be symbolic, as opposed to nonsymbolic stimuli (Raush, 1956); or it may be defined empirically, by means of Luria free association indices (Feffer, 1961). In the face of such diverse operational definitions, it is difficult to maintain a clear referent for "affective stimuli."

Another way of approaching this issue is to regard affective stimuli as being better able to elicit associations from subjects. It is reasonable to assume that human, symbolic, sexual, and aggressive stimuli are more capable of setting off a train of personal and idiosyncratic associations than are more neutral and impersonal stimuli. One way of accounting for the deleterious effect of affective stimuli is to assume that they trigger more associations than do neutral stimuli. This assumption has been verified by Deering (1963), who had subjects associate to affective (pleasant and unpleasant) and neutral words matched for familiarity. Schizophrenics gave significantly more associations to affective words than did normals, but the groups did not differ in the number of associations to neutral words. Presumably, the greater the number of associations to any stimuli, the more the responder will be distracted and the more his performance will suffer.

Insufficient Motivation

Clinical descriptions of schizophrenics usually include apathy and isolation, or at least a tendency to withdraw from interpersonal contacts. The social censure hypothesis attempts to explain the withdrawal (as well as psychological deficit) in terms of the schizophrenic's inordinate sensitivity to rejection. The insufficient motivation hypothesis accepts the disinterest in social contacts as a given and uses it to account for deficit. The schizophrenic is ostensibly unmotivated to perform well on experimental tasks, whereas the normal subject tries to please the experimenter and perseveres at meaningless tasks in the face of boredom and fatigue (Orne, 1962). The schizophrenic, it is suggested, tends to be uncooperative, disinterested, and relatively unmotivated by laboratory criteria of success or by pleasing the experimenter. Supporting this view, Slechta, Gwynn, and Peoples (1963) showed that when such casual social reinforcers as nods of the interviewer's head or "mmm-hm" follow particular verbal behaviors, normal subjects yield a significantly greater percentage of criterion responses than do schizophrenics.

The deficit, then, is not in the schizophrenic's ability but in his motivation. It follows that if the schizophrenic could be urged to cooperate or if he were given appropriate incentives, rewards, or punishments, his performance would equal that of normals.

Cooperation and urging. Wittman (1937) estimated the cooperativeness of schizophrenics, paretics, and nonpsychiatric patients from the response of attendants and physicians to rating scales and checklists. The schizophrenics' cooperativeness scores were generally lower than those of the other two groups, and the schizophrenics were the only group to have high negative correlations between cooperativeness ratings and a memory-reasoning test. Similar findings have been reported by Shakow (1962) for a variety of tasks, and Spohn and Wolk (1963) and Wing and Freudenberg (1961) showed that sustained social stimulation can lead to improved performance and improved ward behavior.

Several experiments were based on the following line of reasoning: (*a*) schizophrenics perform worse than normals, (*b*) schizophrenics are less cooperative or less motivated than normals, (*c*) therefore schizophrenics should improve more than normals when both are urged.

Stotsky (1957) studied the effect of a "motivational incentive" on the psychomotor performance of remitted, regressed schizophrenics and normal subjects. The incentive took the form of encouragement, from their psychotherapist for the patients and from the experimenter for the normal subjects, during an interim between tasks. Both patients and normals were responsive to the motivational condition: the regressed schizophrenics showed significantly greater improvement than normals in simple reaction time (RT), but the normals showed significantly greater improvement than the patients on the disjunctive RT task. D'Alessio and Spence (1963) reported no difference in degree of improvement on a simple motor task between normal and chronic schizophrenic patients. Benton, Jentsch, and Wahler (1960) found that both normal and schizophrenic subjects improved RT performance equally when administered positive "motivating" instructions, and the improvement of the schizophrenic subjects was not significantly different from that of a group of similarly-treated brain-damaged patients (whose data were available from Blackburn, 1958). Olson (1958) reported that schizophrenics improved their performance on a digit-symbol task following either depreciation or praise of a previous performance, whereas normals improved only after praise. However, Goodstein, Guertin, and Blackburn (1961) found no difference in the RT improvement of schizophrenics and normals, both groups improving most following failure instructions. Shankweiler (1959) reported that like schizophrenics, brain-damaged subjects also improve on a psychomotor task subsequent to failure instructions.

Johannsen (1962) had nonparanoid and paranoid schizophrenics and normals cancel the letter f in prose material. At the end of each 2-minute trial subjects were told their performance was good or poor, or they were told nothing. The criticized subjects showed the greatest improvement, the praised subjects were next, and the control subjects were last; these results held for both schizophrenic and normal subjects. The order of improvement for diagnostic groups was: normals, paranoids, and nonparanoids.

Schooler and Spohn (1960) studied schizophrenics and normals in an Asch situation and found both groups equally responsive to social pressure. Felice (1961) examined the effects of interpersonal (experimenter present) and impersonal (experimenter absent) administration on the performance of normals and schizophrenics on a variety of laboratory tasks. Task, diagnosis, and method of administration interacted to produce highly variable results. On some tests schizophrenic performance was depressed by the presence of the experimenter, on other tasks it made no difference, and on some the normals did poorest during the interpersonal administration. However, in these experiments the experimenter's status as a social or affective stimulus may not be the relevant variable. He may simply be an additional distracting element or a source of reduced cues which aid the subject in his task.

In summary, the data on both positively and negatively oriented urging do not support a motivational interpretation of deficit. Normals, schizophrenics, and brain-damaged patients all tend to respond with equal improvement on these conditions. At least one researcher has reported urging instructions to be completely ineffective in modifying schizophrenic psychomoter behavior (Ladd, 1960). In any event, the schizophrenic's deficit has never been completely overcome by these methods. Only Stotsky's (1957) "remitted" group could not be differentiated from normals, and this occurred on some tasks before additional motivations were introduced.

Nonverbal reward and punishment. While the schizophrenic may be insensitive to the rewards of social interaction (pleasing the experimenter, etc.), he might still be affected by more tangible or biologically pertinent reinforcers. Thus an attempt may be made to eliminate his performance deficit by using such nonverbal rewards (cigarettes, money, pleasant stimuli) or punishments (electric shock and "white noise").

Peters (1953) and Peters and Jenkins (1954) studied the motor learning and problem-solving behavior of schizophrenics receiving subshock insulin, providing a "sweet nutrient" as a reward for correct responses. On the basis of his first study, Peters concluded that although the reward did increase operant rate, it did little to eliminate errors. In the second study, using the same incentive condition, an attempt was made to use rewarded problem-solving as an adjunct to therapy. It was anticipated that rewarded achievement on simple motor tasks would provide a base for success in more complex social tasks, and hence would transfer to routine, day to day behavior. It was found that the experimental group had significantly fewer "negative ward incidents" than the control group, but the ratings of ward personnel did not distinguish between the groups. The weak effect found in these studies is consistent with Michaux's (1955) study of apperception as a function of hunger. He reported that unlike normals, schizophrenics show no increase in food completions on a verbal task under hunger conditions.

Schizophrenics, brain-damaged patients, and medical controls were promised money for better performance in a study by Burday (1962). The task involved conceptual sorting, and the money incentive yielded no significant differences among the three groups. However, Rodnick and Garmezy (1957) reported informally that candy, cigarettes, and an attractive scoreboard raise the performance level of schizophrenic subjects. At the beginning of their experiments the performance of good premorbid patients has frequently been no different from that of normal subjects.

Lindsley and Skinner (1954) found results similar to those for dogs, rats, and pigeons, in schizophrenic patients conditioned to pull levers for candy, cigarettes, or projected colored slides. They concluded that psychotic behavior is controlled "to some extent" by the reinforcing properties of the immediate physical environment. Isaacs, Thomas, and Goldiamond (1960) reinstated verbal behavior in two catatonic patients, using operant conditioning methods, and Tilton (1956) reported an improved "general level of functioning" in schizophrenics treated by instrumental learning. However, in Lindsley's study (1960) schizophrenics sometimes abruptly ceased their operant responding without the introduction of any extinction procedure. Furthermore, King, David, and Lovinger (1957) failed to confirm the hypotheses that severity of neuropsychiatric illness is inversely related to rewarded operant rate and that clinical improvement is positively related to this variable. Contrary to Lindsley and Skinner, they concluded that "operant motor behavior seems best classified as a peripheral variable in terms of psychopathology [1957, p. 325]."

As with urging instructions, tangible positive rewards appear somewhat to ameliorate the poor performance of schizophrenics, and less severely disturbed patients may approach the performance of normals on some tasks. However, there is no clear evidence that for schizophrenics, positive tangible rewards are superior to verbal encourgement. Killberg (1962) found that improvement in a verbal learning task was greater for verbally reinforced normal subjects than for similarly treated schizophrenics and that schizophrenics and normals did not differ in improvement when cigarettes were used as a reinforcer. However, the difference between the schizophrenic subjects under the two conditions of reinforcement was not significant.

Whereas monetary reward improved the performance of normals on a serial anticipation task, Topping and O'Connor (1960) found that nonparanoids did not improve and paranoid schizophrenics worsened under the same conditions. Clearly, positive rewards do not exert any consistent control over schizophrenic behavior and probably vary considerably in their consequences with changes in population and experimental setting.

Physical punishment has proved to be the most consistent and effective reinforcer of behavior in schizophrenic subjects. Pascal and Swenson (1952) found no difference between a group of schizophrenic and normal subjects on initial trials of a complex, discrimination RT task, but on subsequent trials the normals improved and the schizophrenics did not. At this point a strong, aversive stimulus was introduced in the form of high intensity noise delivered through earphones, and subjects were told that the noise would be turned off only after a correct response. Under these conditions both groups improved, the schizophrenics showing a greater reduction in latency scores than the normal subjects. Analysis of the final trials revealed no significant difference between the schizophrenic and normal groups.

Cohen (1956) evaluated the performance of chronic schizophrenic and normal subjects on an apparatus which presented patterned visual stimuli. Subjects responded by moving a handle either left or right, depending on the pattern. After a number of trials under the usual conditions of administration, half of the normal group and half of the schizophrenics received a shock at stimulus onset, which was terminated by a correct response. The schizophrenics consistently required more trials to learn and made more errors than did the normals. However, although the unshocked schizophrenics showed a decrement in performance on later trials, the shocked schizophrenics maintained their previous level of performance. This difference between control and experimental groups was not present in the results of the normals.

A study of RT in schizophrenia by Rosenbaum, Mackavey, and Grisell (1957a) suggests that schizophrenics are able to improve their performance under shock conditions. Male and female schizophrenics and hospital employees were first tested in a simple RT situation under standard conditions. Half of each group was then tested under shock conditions, while the remainder continued responding in the previous manner. The task was a simple finger lift from a key in response to an auditory stimulus. In the shock condition subjects were stimulated electrically in the finger to be withdrawn, simultaneously with the sounding of the buzzer. The hospital employees were significantly faster than the psychotics under all conditions of the experiment, but the schizophrenics showed significantly greater improvement than the normals from nonshock to shock conditions. In fact, the electric shock raised the scores of the schizophrenic men to the level of the normals. These results cannot be attributed to an increased "need" to respond quickly because the shock, which was administered to the responding finger, actually replaced sound as the discriminative stimulus. The improvement can as easily be assigned to an increased stimulus dynamism.

In a subsequent study Rosenbaum, Grisell, and Mackavey (1957b) tested the RTs of female schizophrenics and college students. There were several conditions, involving particular combinations of "ready" interval (regular or irregular) and shock. The conditions were called "social motivation," "anxiety motivation," and "biological motivation," respectively. The authors report that:

> The normals were superior to the schizophrenics on all measures, but could be significantly differentiated from the younger, privileged group only on anxiety reaction time [during the anxiety reaction time trials, the shock was presented in an unpredictable manner] [p. 206].

Two other studies are particularly worthy of mention in that a concept formation rather than a psychomotor test provided the experimental task. Cavanaugh (1958) compared the performance of chronic schizophrenics and nonpsychiatric patients and found that under the usual conditions of administration, schizophrenics were inferior. However, when an aversive stimulus was introduced (high intensity noise presented with the concept materials and terminated by a correct response), schizophrenic experimental subjects performed significantly better (in terms of the number of correct concepts and latency scores) than a control (no noise) group. Furthermore, there was no difference in performance between this schizophrenic noise group and a group of similarly treated nonpsychiatric patients. Brown (1961) substantially confirmed Cavanaugh's findings in a study in which incorrect responses were followed by aversive noise.

In brief, in all studies employing physical punishment as an incentive there has been a definite reduction, and in a few instances the temporary elimination, of psychological deficit. These results are consistent with this variant of motivation theory: deficit is a consequence of lowered social motivation, and thus may be reduced when biologically intense reinforcers are used. However, they can also be explained by interference theory: the important fact in the introduction of intense stimuli may be the change in attention and set, with a sharper differentiation between relevant and irrelevant inputs being created.

Verbal reward and punishment. Many studies of schizophrenia have been designed to evaluate the effects of specific verbal cues or reinforcers on the course of psychological tasks. This research has been less concerned with overcoming deficit than with the differential effects of verbal reward and punishment. It will be recalled that Rodnick and Garmezy (1957) hold that social censure or disapproval disrupts the performance of schizophrenics. However, studies of physical punishment and studies which found improvement after negatively oriented urging suggest that the schizophrenic's performance is specifically facilitated by censure.

In a series of experiments Buss and colleagues (Buss & Buss, 1956; Buss, Braden, Orgel, & Buss, 1956; Buss, Wiener, & Buss, 1954) found that the word WRONG, following an incorrect response, yields more effective learning than RIGHT for correct answers. The subjects were mixed groups of psychiatric patients, the majority being schizophrenic. Leventhal (1959) confirmed these findings in a study of diagnosed schizophrenics; "Not so good" was more effective in facilitating verbal conditioning than "Good." Atkinson and Robinson (1961) and Koppenhaver (1961) compared verbal reward and punishment in a paired-associates learning task and found that punishment improved schizophrenic learning, relative to that obtained with reward.

Losen (1961) matched good premorbid schizophrenics with normals on the basis of arithmetic ability and tested them on arithmetic and short-term memory tasks. There were four experimental groups. One group was shown their scores without comment following each response; the other three groups were administered 100%, 50%, and 0% censure schedules. The censure was verbal: "No that was wrong." None of the experimental conditions significantly altered the performance of normal subjects. For schizophrenics, the 0% censure group and the group that was simply shown their scores did not change, but both the 100% and 50% censure groups showed significant improvement. In all of the above studies mild verbal censure of the type used by Garmezy and Rodnick and their students *facilitated* the performance of schizophrenics.

A number of experiments suggest that the ameliorative effects of censure cannot be attributed to the personal or social character of these reinforcers. Lang (1959) found that chronic schizophrenics significantly improved RT scores when a neutral tone followed responses slower than a previously established standard. A subsequent study (Cavanaugh, Cohen, & Lang, 1960) showed that verbal reward was ineffective in improving RT but that verbal censure and the same information about errors used in the previous experiment led to nearly equal improvement. Atkinson and Robinson (1961) found no difference between the facilitating effect of verbal censure and the results obtained when incorrect responses were followed by the sound of an adding machine in operation. On the other hand, Maginley (1956) employed a visual rather than an auditory cue as a nonsocial reinforcer and found that schizophrenics performed less well than normals on the later trials of a concept formation task. However, he attributed this effect to the "diminished interest and attention" occasioned by the light. Johannsen (1961) studied the performance of normals and two schizophrenic groups on a double alternation task. The reinforcers were either verbal right and wrong or diffentially colored light. Paranoid schizophrenics were no different under conditions of social and nonsocial reinforcement but the performance of nonparanoid schizophrenics was significantly poorer when social reinforcers were employed.

The thesis that punishment invariably disrupts performance is clearly not tenable. Both a negative evaluation and specific verbal or physical punishment for errors can lead to significant improvement in performance rather than further deficit. The fact that this improvement occurs in both the presence and absence of socially or personally punitive conditions suggests that the significant factor is information about inadequate responses rather than the interpersonal context.

Knowledge of results is important to any task in which improvement is expected with practice, and in general the normal subject recognizes the correctness or wrongness of a response as soon as it occurs, and no assistance is required. However, schizophrenics seem to be less able to instruct themselves and less able to maintain or usefully alter a response set (Shakow, 1962). Furthermore, studies of incidental learning (Greenberg, 1954; Winer, 1954) reveal that relative to normals, schizophrenics fail to observe objects or relationships towards which their attention has not been specifically directed. Thus, informational cues introduced by the experimenter have greater importance for the psychotic subject in certain tasks.

In the experiments of Lang (1959), Cavanaugh et al. (1960), and Losen (1961) normal subjects performed asymptote under the usual conditions of administration; additional cues did not lead to improvement. Starting at a lower performance level, schizophrenics were able to profit from information about errors and thereby improve their performance. On a task with more ceiling, however, normal subjects can accomplish the same result, Buss and Buss (1956) demonstrated that in concept formation, normals yield a reward-punishment hierarchy no different from that originally found with patients, i.e., punishing errors is more effective in improving performance than rewarding correct responses.

Studies of normal subjects indicate there is considerable difference in the effectiveness of reward and punishment, depending on the experimental task (Meyer & Offenbach, 1962). If the task is a simple one (few alternate responses) and the correct response is already in the respondent's behavioral repertoire, reward is likely to be effective in maintaining the wanted behavior. However, if the task is complex (many alternate responses) and if the correct response is not currently in the subject's repertoire or readily elicited, punishment will provide the better guide. In essence, verbal approval confirms the aptness of a response already made; punishment signals a need for change. In nearly all the experiments under consideration the question raised was to what extent could subjects improve or change, and information about errors generally proved to be the most valuable experimental manipulandum for *both* normal and schizophrenic subjects.

Nevertheless, diagnosis and the specific experimental task interact, and results at variance with this trend are not surprising. Both Atkinson and Robinson (1961) and Leventhal (1959) report that schizophrenics learn fastest when punished for errors, while nonpsychotics show most rapid learning with reward. The former authors attribute these findings to the fact that punishment tends to break up the perseverative behavior unique to schizophrenic subjects, an interpretation consistent with the set hypothesis proposed earlier.

Garmezy's study of generalization (1952) can also be interpreted as suggesting that punishing errors is a more effective method for schizophrenic than normal subjects. The subjects were instructed either to pull a lever if a tone was the same as a training stimulus or to push the lever if it was different. Extreme incorrect pull responses were punished, and correct pull responses were rewarded: in effect, the subject was both rewarded and punished for the same behavior. Under these circumstances the incidence of pull responses among schizophrenic subjects dropped to all tones. They simply learned to avoid the word WRONG, rather than to make the discrimination accomplished by the normal subjects. Maginley (1956) reported similar data in a concept formation task. Reward was superior to a combination of punishment for errors and reward for correct responses, and this held for *both* normal and schizophrenic subjects. It is not clear whether the performance difference may be attributed to censure or the different information provided by the two kinds of cues.

Waters (1962) attempted to separate the effects of censure and information about errors in a two-choice learning task. As might be expected, chronic schizophrenics performed best when provided with maximal cues. Paradoxically, the acute patients had fewer errors under the minimal cue condition. Both groups showed poorer performance when social censure was introduced.

Fischer (1963) used both specific criticism for responses and general criticism for performance, as well as specific praise, general praise, and a control. The task was to repeat a series of lights, and the performance measure was response speed. All subjects improved from the first series to the second, and in order of improvement the conditions were criticism (general and specific), praise (general and specific), and the control. An additional finding was that the subjects performed better in the experimenter's absence (comments delivered via earphones) than in his presence.

In neither Waters' nor Fischer's experiment were there normal subjects, but there were normals in a related study by Brooker (1962), who compared task-relevant criticism (WRONG) and praise (RIGHT) with task-irrelevant criticism and praise. The relevant task was the Wisconsin Card Sorting Test, and the irrelevant task was judging a weight every five trials of the card sorting. Normals and schizophrenics both learned better with task-relevant criticism than with task-relevant praise. In the task-irrelevant condition, criticism led to poorer learning than reward for schizophrenics, but there were no differences for normals. These results can be interpreted as supporting both the social censure and interference theories. In light of the complexities that have cropped up in previous research, the separation of censure from information about errors should be part of any experiment on verbal punishment and deficit.

Summary

The research on affect and reinforcement will be summarized under four headings: affective stimuli, information, punishment, and diagnosis.

Affective stimuli. Certain "affective" stimuli disturb schizophrenics enough to produce psychological deficit, but the stimuli are so varied that this term has no precise referent. Concerning affective stimuli, the social censure hypothesis appears to be incorrect. Schizophrenic deficit occurs not only with stimuli connoting censure but also with human, symbolic, and taboo stimuli. A promising hypothesis is that affective stimuli elicit more associations from schizophrenics and that these associations interfere with performance.

Information. It is clear that schizophrenic subjects improve when they are given information about their responses, and they can profit from cues that may be superfluous to normals. Lang (1959) and Losen (1961) have pointed out that schizophrenics, unlike normals, fail to make statements during tasks which suggest self-guidance (e.g., "I got that one wrong"; "I'll have to watch it next time."). It seems likely that external reinforcing cues serve a directive purpose in schizophrenics that normal subjects accomplish for themselves.

The task usually determines whether positive or negative information is more valuable in improving performance. Most tasks have been complex (initiate many alternate responses), and change (improvement), rather than the maintenance of an initially offered response, has been the desired result. If the task also has sufficient ceiling, both normals and schizophrenics profit more from information about errors than they do from knowledge of successes.

Punishment. In some tasks psychotics make more use of punishment than reward, relative to normal subjects, but there is little evidence to suggest this is due to schizophrenics' personal reaction to social censure. Punishment seems to assist the schizophrenic by breaking up perseverative tendencies, whereas reward maintains a previously correct response that is wrong on subsequent trials. This could happen in paired-associate learning, in which a correct response to one stimulus word is wrong for every other stimulus. The process is probably dependent on instructions the subject gives to himself, and it is here that schizophrenics appear to be deficient.

Physical punishment facilitates the performance of schizophrenics, and in fact it is the only contingency that has led to a complete elimination of deficit on some tasks (concept formation). However, it is conceivable that "noise" and shock serve a focusing or arousal function, rather than the traditional role of incentive or reinforcer.

Diagnosis. Response to motivating instructions, social stimuli, and information about performance all seem to vary with diagnosis. These variables are sensitive to personality differences in normal subjects, and perhaps at least an equal effect should be anticipated in schizophrenics. Because of the variety of populations and experimental situations employed, generalizations about the consequences of motivational factors are difficult. However, paranoid-nonparanoid, good premorbid versus poor premorbid, and the acute-chronic typologies have all proved to be experimentally distinct and cannot be ignored in research with schizophrenics.

CONCEPT ATTAINMENT

Schizophrenics show deficit in conceptual tasks, just as they do in tasks involving simple learning, perception, and psychomotor behavior. In fact, studies on concept learning predominate in the older literature on deficit. There are four main hypotheses in this area: loss of abstractness, loss of communication, regression, and interference.

Loss of Abstractness

Goldstein (1946), analogizing from his research with brain-damaged patients, suggested that the fundamental thinking disorder in schizophrenia is the inability to form abstract concepts. He postulates a disturbance of the "abstract attitude," which has the following characteristics:

> 1. To assume a mental set voluntarily.
> 2. To shift voluntarily from one aspect of the situation to another.
> 3. To keep in mind simultaneously various aspects.
> 4. To grasp the essentials of a given whole; to break up a given whole into parts and to isolate them voluntarily.
> 5. To generalize; to abstract common properties; to plan ahead ideationally; to assume an attitude toward the "mere possible," and to think or perform symbolically.
> 6. To detach our ego from the outer world [p. 19].

The ability to assume the abstract attitude develops gradually being preceded by the "concrete attitude," which is merely a response to immediate sense impression. Goldstein's (1959) major hypothesis is that both organic brain-damaged patients and schizophrenics have lost the abstract attitude and can function in thought and language only at the concrete level. He was careful not to conclude that schizophrenia is the same as organic brain damage, but it is clear from his writings that he attributed schizophrenic deficit in conceptual tasks to a loss of function similar to that seen in neurological conditions.

Goldstein cited as evidence for this theory the relatively poor performance of schizophrenics on the Goldstein-Scheerer test (Bolles & Goldstein, 1938; Goldstein & Scheerer, 1941). Similar results were obtained with the Vigotsky blocks, a test of concept formation (Hanfmann & Kasanin, 1937, 1942; Kasanin, 1946). However, Fisher (1950) found that schizophrenics and hysterics were not significantly different in their performance with the Vigotsky blocks. It is difficult to interpret this body of data. The Goldstein-Scheerer test is nonquantitative and requires a rating of concreteness by the experimenter. The Vigotsky test uses a combination time-help score that confounds slowness with poor performance. Furthermore, in many of the studies adequate control groups were absent. Thus these early studies are inconclusive.

Later investigators, using quantitative techniques and adequate control groups, have demonstrated that schizophrenics are *not* abnormally concrete (Chapman & Taylor, 1957; Fey, 1951; Lothrop, 1960; McGaughran, 1954; McGaughran & Moran, 1956, 1957; Rashkis, 1947; White, 1949, and Williams, 1962). They showed that schizophrenics are capable of responding with abstract concepts, but the concepts are often unusual and idiosyncratic. The problem with schizophrenics is therefore not a loss of the abstract attitude but a tendency to verbalize concepts that are deviant and difficult for normals to understand. One way of interpreting these results is to assume an inability to communicate with others.

Loss of Communication

This theory holds that the problem in schizophrenia is not a loss but a basic lack of communication. The schizophrenic's concepts are bizarre and unacceptable to others because he is withdrawn and asocial. The first one to advance this view was Sullivan (1946), who emphasized language problems. Cameron (1946) elaborated it in terms of fantasy:

> Social communication is gradually crowded out by fantasy; and fantasy itself, because of its nonparticipation in and relation to action, becomes in turn less and less influenced by social patterns. The result

is a progressive loss of organized thinking . . . [pp. 51–52].

As McGaughran (1954; McGaughran & Moran, 1956) has indicated loss of communication and loss of abstractness represent two dimensions of concepts: public-private and open-closed. The public-private dimension is the one emphasized by Sullivan and later Cameron (1946). Schizophrenics ostensibly give an idiosyncratic, private basis for the way they sort objects, whereas normals give a common, public basis. The open-closed dimension is the one emphasized by Goldstein and Kasanin. Schizophrenics give a more closed, narrow, stimulus-bound basis for sorting objects, whereas normals give an open, more inclusive, stimulus-free basis for sorting.

McGaughran and Moran (1956) used a quantified object-sorting task to test these two theories. Sorting performance was scored for both the open-closed and public-private dimensions. Schizophrenics gave slightly fewer public bases for sorting and slightly more open bases than matched normals, but neither difference met the usual level of statistical significance. Furthermore, intelligence was found to be related to the public-private dimension: the more intelligent schizophrenics gave more public bases for sorting than did the less intelligent schizophrenics.

Some of these results were corroborated by Payne and Hewlitt (1960), who included in their battery tests of concreteness. While schizophrenics did perform relatively poorly on three measures of concreteness, performance on these measures was determined mainly by intelligence. These authors also noted that many of the unusual responses of schizophrenics tend to be scored as concrete.

Not all investigators have demonstrated conceptual deficit in schizophrenics. Nathan (1962) compared acute and chronic schizophrenics and normals on eight conceptual tasks, which varied on three dichotomous dimensions: formal versus social, verbal versus nonverbal, and abstract versus subsuming. There were no significant differences between schizophrenics in these tasks.

Evidence that schizophrenics are abnormally concrete is again weak. Their concepts do tend to be eccentric and thus difficult for normals to understand. However, this is a property of the concepts, not a defect in the manner in which they are communicated. Furthermore, both dimensions appear to be related to general intelligence, suggesting that evidence for the concreteness or privacy of concepts is confounded with a more general intellectual disability.

Regression to Childish Thinking

Goldman (1962) elaborated Werner's (1948) developmental approach, combining the loss of abstractness and the loss of communication views into a single theory. He distinguished three dimensions believed to be important in the development of thinking in children. First, the child's concepts and ideas are initially personal and idiosyncratic, but they gradually become more public and common. Second, early concepts are labile and shifting, whereas later ones are more stable. Third, early concepts are more concrete and tied to the stimulus context, but later concepts are more abstract and free of the stimuli or events with which they were originally associated.

In support of his position Goldman cited qualitative findings with children and clinical reports on schizophrenics' concreteness (Arieti, 1955; Cameron, 1938; and Kasanin, 1946). Unfortunately these findings lack quantification or adequate controls. As additional evidence Goldman noted that schizophrenics tend to define words more concretely than normals (Choderkoff & Mussen, 1952; Feifel, 1949; Flavell, 1956; Harrington & Ehrmann, 1954). However, as we noted above, bizarre and eccentric responses tend to be scored as concrete, which renders such evidence inconclusive. Furthermore, a number of experiments have demonstrated that schizophrenics give abstract concepts.

Going beyond these considerations, there is the more general issue of the regression hypothesis: does the thinking of schizophrenics resemble that of children? Feifel (1949) showed that children and schizophrenics defined words in a similar pattern that deviated from normal adult definitions. Ellsworth (1951) found that children and schizophrenics were similar in the way they used different parts of speech. Burstein (1959, 1961) demonstrated that children and schizophrenics both tended to equate antonyms with synonyms to a greater extent than normals. On the other hand, Cameron (1938) found that children were very different from schizophrenics in the way they completed incomplete sentences involving causality.

Chapman, Burstein, Day, and Verdone (1961) pointed out a possible artifact in these studies:

> the tasks were such that both children and schizophrenics were characterized by deviant performance relative to normal adults. However, as a result of restrictions either on the response alternatives available or on the investigator's categorization of the responses, each task provided relatively few avenues along which deviations could occur. The tasks tended to measure undifferentiated deviation rather than different kinds of specific deviation. Therefore the likelihood that schizophrenics and children would appear similar was very great [p. 541].

Chapman et al. administered two different thinking tasks to children, schizophrenics, and brain-damaged patients. On one task the younger children resembled schizophrenics but not brain-damaged patients. On the other task the younger children resembled brain-damaged patients but not schizophrenics. They concluded:

> There is no blanket similarity between the error patterns of children and of either schizophrenics or brain-damaged patients, and the use of the term "regression" to imply such a blanket similarity is not justified [p. 545].

To the extent that these comments are true, we may question the conclusion that the thinking of schizophrenics and children is similar. In brief, the evidence for regression of thought processes is equivocal, and there are opposing findings; we regard the regression hypothesis as not proved.

Interference Theory

In searching for an approach to conceptual deficit, one possibility is interference theory. There are two variants, one emphasizing overinclusion or excessive generalization and the other a deficit in attention.

Overinclusion. Overinclusion is the tendency to include irrelevant and extraneous aspects in responding to stimuli. In Cameron's (1938) early work he demonstrated that schizophrenics have difficulty in maintaining boundaries, their concepts spilling over into nontask stimuli such as objects in the room and even the experimenter. Cameron labeled this *interpenetration:* the intrusion of personal and idiosyncratic themes into the schizophrenic's speech and concepts.

Cameron and Magaret (1951) have pointed out that *exclusion* is necessary for success on a given task. At first the subject attends to too many aspects of the stimulus situation, and much of his response is superfluous. Gradually he focuses on the crucial aspects of the stimulus situation, ignoring irrelevant aspects; similarly, he drops out motor responses he does not need. Presumably the schizophrenic cannot exclude nonessential stimulus elements, and his overinclusiveness impairs performance on conceptual tasks.

That schizophrenics are overinclusive has been established by a number of experiments since Cameron's original observations. Zaslow (1950) showed that schizophrenics included more stimuli in the concepts of circularity and triangularity than did normals, although Kugelmass and Fondeur (1955) did not replicate these results with early schizophrenics.

Epstein (1953) and Moran (1953) presented subjects with a stimulus word and a number of response words. The task was to underline response words that were an essential part of the concept denoted by the stimulus. In both studies schizophrenics underlined significantly more response words, especially more distantly related words, than did normals. In addition Moran found that the schizophrenics' associations to the stimulus words were more distant and their synonyms were less precise than those of normals.

Lovibond (1954) quantified object-sorting performance and found that schizophrenics were more overinclusive than normals. Chapman (1956) and Chapman and Taylor (1957) presented pictures of different objects and had subjects sort them under specific headings or concepts such as clothing, furniture, and fruit. In comparison to normals, schizophrenics were more influenced by irrelevant and distracting elements, their conceptual performance being severely impaired by extraneous stimuli.

Payne and his collaborators (Payne & Hewlitt, 1960; Payne, Matussek & George, 1959) used as conceptual tasks a set of proverbs, two sorting tests, and Epstein's word-grouping test. The results were in the expected direction and for the most part significant: schizophrenics were overinclusive in comparison to normals. A number of the schizophrenics were no more overinclusive than the normals, but they were somewhat slow in their motor responses. Payne and Hewlitt speculated that paranoid and catatonic schizophrenics behave differently in these tasks. Paranoids, they suggested, tend to overgeneralize, their overinclusiveness showing up both in ideas of reference and in laboratory conceptual tasks; but they are neither slower nor less abstract than normals. Catatonics do not overgeneralize and are not overinclusive, but they tend to be slower and less abstract than normals.

Attention. Chapman (1961), who reported strong evidence for the overinclusion hypothesis, was not completely satisfied with it. He argued that schizophrenic concreteness is really *underinclusion* (overexclusion), which means that schizophrenics are both overinclusive and underinclusive:

> what may in part underlie the frequently observed inappropriate enlarging or restricting of the application of common concepts is in fact a preference for concepts of a particular breadth [p. 514].

He used two kinds of tasks, one tending to elicit errors of overinclusion and one tending to elicit errors of overexclusion. Schizophrenics were found to make both kinds of errors, although there was a predominance of overinclusion errors. When the task called for broad concepts, there were more overexclusion errors and fewer overinclusion errors. The same tendency was observed in normals, although greatly diminished. Thus the schizophrenics' predilection for concepts of a particular breadth may be merely an exaggeration of a similar finding in normals.

This suggestion is supported by a study by Seth and Beloff (1959) on schizophrenics and tubercular controls. The schizophrenics, who were inferior on verbal tasks, manifested a language decrement and a tendency not to use abstract concepts spontaneously. The errors made by the schizophrenics were exaggerations of the error tendencies of the controls, and the authors suggested that the underlying reason for this was a lowering or alternation of the attention process.

These results lead to the second variant of the interference hypothesis:

> Both disorders of perception and thinking in schizophrenic patients are secondary to a disorder of the span of attention, which can be too broad or too narrow, or may alternate between the two. Constancy of perception . . . depends on the ability to perceive a thing in its context, or to take into account all the "cues" existing in the whole perceptual field. Thus it is related to "broadness of attention." "Overinclusiveness" or wide span of schizophrenic concepts is related also to "overbroadness" of their attention, which makes them incapable of excluding irrelevant stimuli. On the other hand, the ability to perceive embedded figures is related to "narrowness" of attention which causes the distracting, embedded figure to be excluded [Weckowicz & Blewitt, 1959, p. 914].

Note that this hypothesis, introduced before Chapman's results appeared, explains his results. Excessively broad attention should lead to overinclusive concepts; excessively narrow attention should lead to overexclusive concepts; alternating between too-broad and too-short attention should lead to both overinclusive errors and overexclusive errors, such as were found in schizophrenics by Chapman (1961).

Weckowicz and Blewitt were able to test one implication of their hypothesis. In maintaining constancy of perception it is necessary to accept certain cues and reject others. There must be some kind of "filter" mechanism that avoids the errors of both overinclusiveness and overexclusiveness in conceptual performance as well as in size perception. They obtained high correlations (in the fifties) between abstraction scores on a sorting test and size constancy, thus substantiating the hypothesis. More evidence of this kind is needed, but at present this second variant of the interference hypothesis appears to be a promising explanation of schizophrenic deficit in the conceptual area.

Summary

This section has considered four interpretations of schizophrenic deficit in concept attainment. The loss of abstractness theory appears to be incorrect: schizophrenics' concepts are not especially concrete although they tend to be eccentric and deviant. Schizophrenics' concepts are not necessarily childlike, and the evidence for such regression may be attributed to both the limited responses available on the tasks used and to the scoring of idiosyncratic responses as concrete.

Concerning lack of communication, there is no doubt that schizophrenics have difficulty in making their concepts comprehensible to others. However, this appears to be due to the bizarre nature of the concepts rather than an inability to communicate them. Thus the disorder appears to involve thought processes rather than communication skills.

Interference theory appears to be the most promising approach. The overinclusion hypothesis has been sustained by a number of studies: schizophrenics' concepts are excessively broad, and they suffer from the intrusion of extraneous and irrelevant elements. Whether this is true of all schizophrenics or mainly of paranoids is an interesting question. One possibility is that paranoids are more inclusive and less deteriorated than nonparanoids (catatonics and hebephrenics).

DISCUSSION

This section will focus on five theories that attempt to explain schizophrenic deficit in the two research areas covered in this paper. Three theories emphasize interpersonal and social aspects: social censure, sensitivity to affective stimuli, and insufficient motivation. The other two, while not ignoring interpersonal aspects completely, emphasize more impersonal aspects of psychological functioning: regression and interference.

Social Censure

This theory exists in two basic forms, one general and the other a specific corollary. The general theory holds that schizophrenics are abnormally sensitive to, and disrupted by, social censure and stimuli connoting censure. In its support is the fact that in some experiments deficit has been found to increase when censure stimuli are introduced. However, other research has found no difference in the censure response of schizophrenics and normals, or effects for one type of schizophrenic and not for another. Furthermore, data interpreted in support of censure theory are often more parsimoniously explained in other terms (differential information or the saliency of cues). The strongest negative evidence is that on a great variety of tasks social censure *reduces* deficit in schizophrenia rather than increases it, i.e., facilitates performance relative to praise or no reinforcement.

The specific censure theory argues that schizophrenics with good and poor premorbid histories are differentially responsive to parental chastisement: the former are disturbed by paternal and the latter by maternal censure. While there are data consistent with this view, replication has infrequently supported original studies, and methodological problems abound. Both schizophrenic groups have shown disturbance with parental stimuli, regardless of whether the stimuli were positive or negative; both types of schizophrenics have, in specific experiments, been similar to normals in their response to censure.

Furthermore, while distinctions are made within the schizophrenic group, control subjects are treated as a homogeneous population. We cannot help but speculate that the subdivision of a hospital attendant or neurotic group—following similar criteria of general social or sexual adequacy—might yield similar differences in response to parental stimuli. In other words, we do not know that the findings are in any way unique to schizophrenia. In fact, while good and poor premorbid patients are treated as subtypes of schizophrenia, group assignment is actually determined by a scale defining a continuous dimension of social maturity (Phillips, 1953). The scales' author has already provided strong evidence:

> that the relationship of achieved level of maturity (defined in terms of premorbid social competence) to certain dimensions of psychopathology is not unique to schizophrenia, but instead cuts across all forms of functional mental disorder [Zigler & Phillips, 1962, p. 216].

As a general theory of deficit the social censure approach is not sufficiently comprehensive and involves a major inconsistency: stimulus conditions other than censure have been found to yield deficit in schizophrenia, and censure may reduce rather than increase deficit. The specific theory involves a questionable typology. It also needs the support of less equivocal experiments. Dividing schizophrenics into good and poor premorbid groups does not account for the variable effects of censure, a full explanation of which must include *task variables*.

This evaluation does not necessarily negate the clinical theory from which the social censure hypothesis originated. Parental censure may still be an important variable in the histories of schizophrenics and may shape some symptoms. However, social censure does not provide a general explanation for schizophrenic deficit in the laboratory.

Sensitivity to Affective Stimuli

This theory resembles social censure theory, but it is broader in that it assumes that schizophrenics are especially sensitive to, and disrupted by, *all* affective stimuli. This assumption has been verified in many experiments which have demonstrated poorer performance by schizophrenics with affective stimuli than with neutral stimuli. The difficulty with this evidence is that so many stimuli have been labeled *affective* that the term no longer has precise referents. The underlying assumption in this theory is that past events in schizophrenics' lives have rendered them abnormally sensitive to stimuli that connote traumatic situations or anxiety-laden interpersonal relationships. While some of the affective stimuli that lead to deficit pertain to special areas of maladjustment such as sexuality, censure, or competition, others have only a vague connection with personal adjustment or emotion, e.g., human-nonhuman, symbolic-nonsymbolic, and patient-attendant stimulus dimensions have all been shown to influence the performance of schizophrenics. Furthermore, as we have seen with censure stimuli, the effect of specific affective contents varies considerably from experiment to experiment, and perhaps from subject to subject.

The dilemma is as follows. If the term affective is loosely defined, the theory can account for all the data, but it lacks precision. If the term affective is restricted to a narrow, precise meaning, then the reliability of the phenomenon must be questioned; furthermore, nonaffective stimuli have been found to produce deficit. Shakow (1962) has suggested that the associations of schizophrenics constitute a kind of apperceptive mass,

> full of elements—both affective and nonaffective—of past experience. For the schizophrenic, many of these elements are floating around on top of the barrel, ready to be attached to almost any new situation [p. 9].

This conception accommodates the data and provides the basis for an interference explanation. Evidence for this view will be assessed in the next paper. However, the psychodynamic concept of affective stimuli is clearly not a sufficient theory of schizophrenic deficit.

Insufficient Motivation

Another group of theorists argue that schizophrenic deficit is attributable not to an oversensitivity to social stimuli but to a reduced sensitivity to such inputs. Schizophrenics are held to be unmotivated, uninterested in pleasing the experimenter, or uninterested in meeting task requirements; Hunt and Cofer (1944) suggested that responses to social stimuli have been extinguished in schizophrenic patients. In general, the evidence is against this formulation. Testable schizophrenics are as responsive to persuasion as normals or brain-damaged patients whether the social pressure is based on encouragement or on critical admonishment. Many studies stimulated by this theory have confounded the effects of praise and criticism or reward and punishment with the effects of giving subjects different information about their task performance. Evaluating their separate effects is difficult. Nevertheless, schizophrenics and normals generally show a similar pattern of responding: on most tasks information about errors leads to better performance than the signaling of correct responses. However, the importance of this additional information seems to be greater for schizophrenics than for normal subjects. Schizophrenics show relatively greater improvement with clear supplementary cues than do normals, and there is some evidence that information about errors is particularly useful to schizophrenics. The effect of these cues does not appear to be social motivational but rather to help maintain attention and guide responding. It has been suggested that the schizophrenic fails to instruct himself as normals do, and the additional information provided after each response fulfills this guidance function.

Physical punishment tends to facilitate the performance of schizophrenics, and it has occasionally eliminated schizophrenic deficit. This fact has been used to support the lowered motivation hypothesis: if schizophrenics are able to approach normal functioning when reinforced with pain termination but not when social reinforcers are administered, a deficit in social motivation is implied. An alternate view is that physical punishment may be a better source of information about errors, breaking up incorrect sets, and guiding responding. The importance of cue emphasis as opposed to motivational instigation has been demonstrated in studies of normal subjects, in which improvement was occasioned by delivering electric shock for correct responses. Furthermore, when intense stimuli are coincident with important discriminative stimuli, the latter are lent emphasis and more clearly separated from the irrelevant cues in the situation. Evidence for these alternate motivation and interference interpretations of studies employing aversive stimuli will be evaluated in the subsequent paper.

Regression

The loss of abstractness theory received support mainly from early studies, but later work revealed that the excessive concreteness of schizophrenics may be attributed to methods of scoring responses rather than to a loss of the abstract attitude. Schizophrenics are capable of attaining abstract concepts, although their concepts may be bizarre, eccentric, and deviant from those given by normals.

The loss of abstractness theory is but one variant of regression theory, which includes not only Goldstein's formulation but those of Freud, Arieti (1955), and Werner (1948). Regression theory makes two fundamental assumptions: (*a*) there is a fixed sequence of developmental stages that ends in maturity or adult normality, and (*b*) psychopathology represents a retracing of these developmental steps. Concerning deficit, the theory should demonstrate: (*a*) a fixed sequence or hierarchy of psychomotor learning and thinking behavior in childhood, and (*b*) a retracing of the sequence in psychopathology. The only one to attempt such a specific formulation with respect to schizophrenic deficit has been Goldman (1962), using Werner's developmental approach. Whether or not one accepts Goldman's developmental sequence of learning and thinking, there is considerable doubt that the learning and thinking of schizophrenics resembles those of children of any given age. There are occasional similarities between schizophrenics' concepts and those of children: tendencies toward more primitive concepts and deviance from adult concepts. However, there are also marked differences between children's and schizophrenics' concepts: those of schizophrenics are usually more abstract, bizarre, and eccentric than those of children.

Furthermore, it is not sufficient to demonstrate vague similarities between schizophrenics and children. It must be shown that as an individual becomes schizophrenic, he retraces the developmental sequence (if there is one) of learning and thinking; as he recovers, he again moves forward toward demonstrably more mature modes of learning and thinking. These corollaries of regression theory have yet to be established empirically, and therefore we conclude that the regression theory of schizophrenic deficit is unproved.

Interference

This theory assumes that when a schizophrenic is faced with a task, he cannot attend properly or in a sustained fashion, maintain a set, or change the set quickly when necessary. His ongoing response tendencies suffer interference from irrelevant, external cues and from "internal" stimuli which consist of deviant thoughts and associations. These irrelevant, distracting, mediated stimuli prevent him from maintaining a clear focus on the task at hand, and the result is psychological deficit.

If the schizophrenic has difficulty in shifting to a new set, he should benefit from stimuli that break the old set. This is precisely what has been established. Schizophrenics benefit from punishment, which eliminates previously correct response tendencies that are no longer appropriate. Physical punishment, in some instances, has been found to eliminate schizophrenic deficit entirely.

If the presence of distracting, internal stimuli prevents the schizophrenic from giving self-instructions (as normals do), he should benefit from external cues. It has been established that schizophrenics are helped more than normals by external reinforcing cues which help direct responses, although normals are also helped. Thus interference theory has an explanation for the effects of punishment and of the rewarding and punishing stimuli that follow responses.

Interference theory can account for schizophrenic deficit with affective stimuli. The theory assumes that the schizophrenics' associations distract him, thereby producing disturbance of performance. The more associations, the greater the deficit. Since affective stimuli elicit more associations than neutral stimuli (Deering, 1963), it follows that they should produce more schizophrenic deficit.

Conceptual deficit is explained by assuming excessive variability in attention. The schizophrenic is either overinclusive because he allows irrelevant stimuli to intrude or overexclusive because he attempts to defend against the distracting, internal stimuli. It follows that some schizophrenics should be overinclusive, some overexclusive, and some alternating between the two; all three possibilities have been found to exist in schizophrenics when they are confronted with conceptual tasks.

Interference theory may be viewed as a devil's advocate in that in each research area it offers an alternative to the theories of social censure, sensitivity to affective stimuli, insufficient motivation, and regression. Obviously, the theory is sufficiently comprehensive, but a theory must do more than merely offer alternatives. Also needed is evidence that tests specific hypotheses, in this instance about the particular interfering effects of environmental, associational, and also somatic stimuli. Such evidence is reviewed in the subsequent paper.

REFERENCES

Alvarez, R. R. A comparison of the preferences of schizophrenic and normal subjects for rewarded and punished stimuli. Unpublished doctoral dissertation, Duke University, 1957.

Arey, L. B. The indirect representation of sexual stimuli by schizophrenic and normal subjects. *Journal of Abnormal and Social Psychology*, 1960, **61**, 424–431.

Arieti, S. *Interpretation of schizophrenia.* New York: Bruner, 1955.

Atkinson, Rita L., & Robinson, Nancy M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, **62**, 322–326.

Baxter, J. C., & Becker, J. Anxiety and avoidance behavior in schizophrenics in response to parental figures. *Journal of Abnormal and Social Psychology*, 1962, **64**, 432–437.

Benton, A. L., Jentsch, R. C., & Wahler, H. J. Effects of motivating instructions on reaction time in schizophrenia. *Journal of Nervous and Mental Disease*, 1960, **130**, 26–29.

Blackburn, H. L. Effects of motivating instructions on reaction time in cerebral disease. *Journal of Abnormal and Social Psychology*, 1958, **56**, 359–366.

Bleke, R. C. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *Journal of Personality*, 1955, **23**, 479–498.

Bolles, Marjorie, & Goldstein, K. A study of schizophrenic patients. *Psychiatric Quarterly*, 1938, **12**, 42–65.

Brodsky, M. J. Interpersonal stimuli as interference in a sorting task. *Dissertation Abstracts*, 1961, **22**, 2068.

Brooker, H. The effects of differential verbal reinforcement on schizophrenic and non-schizophrenic hospital patients. Unpublished doctoral dissertation, Indiana University, 1962.

Brown, R. L. The effects of aversive stimulation on certain conceptual error responses of schizophrenics. *Dissertation Abstracts*, 1961, **22**, 629.

Burday, G. The performance of schizophrenic, brain-damaged and nonpsychiatric patients on a modified matching concept formation test and the effects of positive motivation on concept formation performance. Unpublished doctoral dissertation, Temple University, 1962.

Burstein, A. G. Primary process in children as a function of age. *Journal of Abnormal and Social Psychology*, 1959, **59**, 284–286.

Burstein, A. G. Some verbal aspects of primary process thought in schizophrenia. *Journal of Abnormal and Social Psychology*, 1961, **62**, 155–157.

Buss, A. H., Braden, W., Orgel, A., & Buss, Edith H. Acquisition and extinction with different verbal reinforcement combinations. *Journal of Experimental Psychology*, 1956, **52**, 288–295.

Buss, A. H., & Buss, Edith H. The effect of verbal reinforcement combinations on conceptual learning. *Journal of Experimental Psychology*, 1956, **52**, 283–287.

Buss, A. H., Wiener, M., & Buss, Edith H. Stimulus generalization as a function of verbal reinforcement combinations. *Journal of Experimental Psychology*, 1954, **48**, 433–436.

Cameron, N. S. Reasoning, regression and communication in schizophrenics. *Psychological Monographs*, 1938, **50**(1, Whole No. 221).

Cameron, N. S. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1946. Pp. 50–64.

Cameron, N., & Magaret, Ann. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

Cavanaugh, D. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *Journal of Abnormal and Social Psychology*, 1958, **57**, 8–12.

A 477

CAVANAUGH, D., COHEN, W., & LANG, P. J. The effect of "social censure" and "social approval" on the psychomotor performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1960, **60**, 213–218.

CHAPMAN, L. J. The role of type of distracter in the "concrete" performance of schizophrenics. *Journal of Personality*, 1956, **25**, 130–141.

CHAPMAN, L. J. A reinterpretation of some pathological disturbances in conceptual breadth. *Journal of Abnormal and Social Psychology*, 1961, **62**, 514–519.

CHAPMAN, L. J., BURSTEIN, A. G., DAY, DOROTHY, & VERDONE, P. Regression and disorders of thought. *Journal of Abnormal and Social Psychology*, 1961, **63**, 540–545.

CHAPMAN, L. J., & TAYLOR, JANET A. The breadth of deviate concepts used by schizophrenics. *Journal of Abnormal and Social Psychology*, 1957, **54**, 118–123.

CHODERKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *Journal of Consulting Psychology*, 1952, **16**, 43–48.

COHEN, B. D. Motivation and performance in schizophrenia. *Journal of Abnormal and Social Psychology*, 1956, **52**, 186–190.

CULVER, C. M. The effect of cue value on size estimation in schizoprenic subjects. Unpublished doctoral dissertation, Duke University, 1961.

D'ALESSIO, G. R., & SPENCE, JANET T. Schizophrenic deficit and its relation to social motivation. *Journal of Abnormal and Social Psychology*, 1963, **66**, 390–393.

DAVIS, R. H., & HARRINGTON, R. W. The effect of stimulus class on the problem-solving behavior of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1957, **54**, 126–128.

DEERING, GAYLE. Affective stimuli and disturbance of thought processes. *Journal of Consulting Psychology*, 1963, **27**, 338–343.

DEWOLFE, A. S. The effect of affective tone on the verbal behavior of process and reactive schizophrenics. *Journal of Abnormal and Social Psychology*, 1962, **64**, 450–455.

DUNHAM, R. M. Sensitivity of schizophrenics to parental censure. Unpublished doctoral dissertation, Duke University, 1959.

DUNN, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *Journal of Personality*, 1954, **23**, 48–64.

EHRENWORTH, J. The differential responses to affective and neutral stimuli in the visual-motor performance of schizophrenics and normals. Unpublished doctoral dissertation, Boston University, 1960.

ELLSWORTH, R. B. The regression of schizophrenic language. *Journal of Consulting Psychology*, 1951, **15**, 378–391.

EPSTEIN, S. Overinclusive thinking in a schizophrenic and a control group. *Journal of Consulting Psychology*, 1953, **17**, 384–388.

FEFFER, M. H. The influence of affective factors on conceptualization in schizophrenia. *Journal of Abnormal and Social Psychology*, 1961, **63**, 588–596.

FEIFEL, H. Qualitative differences in the vocabulary response of normals and abnormals. *Genetic Psychology Monographs*, 1949, **39**, 151–204.

FELDSTEIN, S. The relationship of interpersonal involvement and affectiveness of content to the verbal communication of schizophrenic patients. *Journal of Abnormal and Social Psychology*, 1962, **64**, 39–45.

FELICE, A. Some effects of subject-examiner interaction on the task performance of schizophrenics. Unpublished doctoral dissertation, Temple University, 1961.

FEY, ELIZABETH T. The performance of young schizophrenics on the Wisconsin Card Sorting Test. *Journal of Consulting Psychology*, 1951, **15**, 311–319.

FISCHER, E. H. Task performance of chronic schizophrenics as a function of verbal evaluation and social proximity. *Journal of Clinical Psychology*, 1963, **19**, 176–178.

FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychological Monographs*, 1950, **64**(1, Whole No. 307).

FLAVELL, J. H. Abstract thinking and social behavior in schizophrenia. *Journal of Abnormal and Social Psychology*, 1956, **52**, 208–211.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality*, 1952, **21**, 253–276.

GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease*, 1959, **129**, 450–466.

GOLDMAN, A. E. A comparative-developmental approach to schizophrenia. *Psychological Bulletin*, 1962, **59**, 57–69.

GOLDSTEIN, K. Methodological approach to the study of schizophrenic thought disorder. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1946. Pp. 17–40.

GOLDSTEIN, K. Concerning the concreteness in schizophrenia. *Journal of Abnormal and Social Psychology*, 1959, **59**, 146–148.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychological Monographs*, 1941, **53**(2, Whole No. 239).

GOODSTEIN, L. D., GUERTIN, W. H., & BLACKBURN, H. L. Effects of social motivational variables on choice reaction time in schizophrenics. *Journal of Abnormal and Social Psychology*, 1961, **62**, 24–27.

GREENBERG, A. Directed and undirected learning in chronic schizophrenia. *Dissertation Abstracts*, 1954, **14**, 1457–1458.

HANFMANN, EUGENIA, & KASANIN, J. A method for the study of concept formation. *Journal of Psychology*, 1937, **3**, 521–540.

HANFMANN, EUGENIA, & KASANIN, J. S. Conceptual thinking in schizophrenia. *Nervous and Mental Disease Monographs*, 1942, No. 67.

HARRINGTON, R., & EHRMANN, J. C. Complexity of response as a factor in the vocabulary performance of schizophrenics. *Journal of Abnormal and Social Psychology*, 1954, **49**, 362–364.

Harris, J. G., Jr. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *Journal of Personality,* 1957, **25,** 651–671.

Hunt, J. McV., & Cofer, C. Psychological deficit in schizophrenia. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 2. New York: Ronald Press, 1944. Pp. 971–1032.

Isaacs, W., Thomas, J., & Goldiamond, I. Application of operant conditioning to reinstate verbal behavior in psychotics. *Journal of Speech and Hearing Disorders,* 1960, **25,** 8–12.

Johannsen, W. J. Responsiveness of chronic schizophrenics and normals to social and nonsocial feedback. *Journal of Abnormal and Social Psychology,* 1961, **62,** 106–113.

Johannsen, W. J. Effect of reward and punishment on motor learning by chronic schizophrenics and normals. *Journal of Clinical Psychology,* 1962, **18,** 204–207.

Kasanin, J. S. The disturbance of conceptual thinking in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1946. Pp. 41–49.

Killberg, J. The differentiating effects of nonverbal and verbal rewards in the modification of verbal behavior of schizophrenic and normal subjects. Unpublished doctoral dissertation, Columbia University, 1962.

King, G., David, M., & Lovinger, E. Operant motor behavior in acute schizophrenics. *Journal of Personality,* 1957, **25,** 317–326.

Koppenhaver, N. D. The effects of verbal and nonverbal reinforcement on the performance of schizophrenic subjects. Unpublished doctoral dissertation, Purdue University, 1961.

Kreinik, Phyllis S. Parent-child themas and concept attainment in schizophrenia. Unpublished doctoral dissertation, Duke University, 1959.

Kugelmass, S., & Foundeur, M. P. Zaslow's test of concept formation: Reliability and validity. *Journal of Consulting Psychology,* 1955, **19,** 227–229.

Ladd, C. E. The digit symbol performance of schizophrenic and nonpsychiatric patients as a function of motivational instructions and task difficulty. Unpublished doctoral dissertation, University of Iowa, 1960.

Lang, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *Journal of Abnormal and Social Psychology,* 1959, **59,** 263–268.

Lebow, K. E., & Epstein, S. Thematic and cognitive responses of good premorbid schizophrenics to cues of nurturance and rejection. *Journal of Consulting Psychology,* 1963, **27,** 24–33.

Leventhal, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *Journal of Abnormal and Social Psychology,* 1959, **59,** 162–166.

Lindsley, O. Characteristics of the behavior of chronic psychotics as revealed by free-operant conditioning methods. *Diseases of the Nervous System Monograph Supplement,* 1960, **21.**

Lindsley, O., & Skinner, B. F. A method for the experimental analysis of the behavior of psychotic patients. *American Psychologist,* 1954, **9,** 419–420.

Losen, S. M. The differential effect of censure on the problem solving behavior of schizophrenics and normal subjects. *Journal of Personality,* 1961, **29,** 258–272.

Lothrop, W. W. Psychological test covariates of conceptual deficit in schizophrenia. *Journal of Consulting Psychology,* 1960, **24,** 496–499.

Lovibond, S. H. The Object Sorting Test and conceptual deficit in schizophrenia. *Australian Journal of Psychology,* 1954, **6,** 52–70.

McGaughran, L. S. Predicting language behavior from object sorting. *Journal of Abnormal and Social Psychology,* 1954, **49,** 183–195.

McGaughran, L. S., & Moran, L. J. "Conceptual level" vs. "conceptual area" analysis of object sorting behavior of schizophrenic and nonpsychiatric groups. *Journal of Abnormal and Social Psychology,* 1956, **52,** 43–50.

McGaughran, L. S., & Moran, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *Journal of Abnormal and Social Psychology,* 1957, **54,** 44–49.

Maginley, H. J. The effect of "threats" of failure upon the conceptual learning performance of hospitalized mental patients. Unpublished doctoral dissertation, University of Pittsburgh, 1956.

Marx, A. The effect of interpersonal content on conceptual task performance of schizophrenics. Unpublished doctoral dissertation, University of Oklahoma, 1962.

Meyer, W. J., & Offenbach, S. I. Effectiveness of reward and punishment as a function of task complexity. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 532–534.

Michaux, W. Schizophrenic apperception as a function of hunger. *Journal of Abnormal and Social Psychology,* 1955, **50,** 53–58.

Moran, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychological Monographs,* 1953, **67**(20, Whole No. 370).

Moriarty, D., & Kates, S. L. Concept attainment on materials involving social approval and disapproval. *Journal of Abnormal and Social Psychology,* 1962, **65,** 355–364.

Nathan, P. A comparative investigation of conceptual ability in relation to frustration tolerance. Unpublished doctoral dissertation, Washington University, 1962.

Neiditch, S. J. Differential response to failure in hospital and nonhospital groups. *Journal of Abnormal and Social Psychology,* 1963, **66,** 449–453.

Nelson, Sandra, & Caldwell, W. E. Perception of affective stimuli by normal and schizophrenic subjects in a depth perception task. *Journal of General Psychology,* 1962, **67,** 323–335.

Olson, G. W. Failure and subsequent performance of schizophrenics. *Journal of Abnormal and Social Psychology,* 1958, **57,** 310–314.

Orne, M. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17,** 776–783.

Pascal, C., & Swensen, G. Learning in mentally ill patients under unusual motivation. *Journal of Personality,* 1952, **21,** 240–249.

Payne, R. W., & Hewlett, J. H. G. Thought disorder in psychotic patients. In H. J. Eysenck

(Ed.), *Experiments in personality*. Vol. II. London: Routledge & Kegan Paul, 1960. Pp. 3–104.

Payne, R. W., Mattussek, P., & George, E. I. An experimental study of schizophrenic thought disorder. *Journal of Mental Science,* 1959, **105**, 627–652.

Peters, H. N. Multiple choice learning in the chronic schizophrenic. *Journal of Clinical Psychology,* 1953, **9**, 328–333.

Peters, H. N., & Jenkins, R. L. Improvement of chronic schizophrenic patients with guided problem-solving motivated by hunger. *Psychiatric Quarterly,* 1954, **28**, 84–101.

Phillips, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease,* 1953, **117**, 515–525.

Pishkin, V., Smith, T. E., & Leibowitz, H. W. The influence of symbolic stimulus value on perceived size in chronic schizophrenia. *Journal of Consulting Psychology,* 1962, **26**, 323–330.

Rashkis, H. A. Three types of thinking disorder. *Journal of Nervous and Mental Disease,* 1947, **106**, 650–670.

Raush, H. L. Object constancy in schizophrenia: The enchancement of symbolic objects and conceptual stability. *Journal of Abnormal and Social Psychology,* 1956, **52**, 231–234.

Rodnick, E. H., & Garmezy, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

Rosenbaum, G., Mackavey, W. R., & Grisell, J. L. Effects of biological and social motivation on schizophrenic reaction time. *Journal of Abnormal and Social Psychology,* 1957, **54**, 364–368. (a)

Rosenbaum, G., Grisell, J. L., & Mackavey, W. R. The relationship of age and privilege status to reaction time indices of schizophrenic motivation. *Journal of Abnormal and Social Psychology,* 1957, **55**, 202–207. (b)

Schooler, C., & Spohn, H. E. The susceptibility of chronic schizophrenics to social influence in the formation of perceptual judgments. *Journal of Abnormal and Social Psychology,* 1960, **61**, 348–354.

Seth, G., & Beloff, Halla. Language impairment in a group of schizophrenics. *British Journal of Medical Psychology,* 1959, **32**, 288–293.

Shakow, D. Segmental set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry,* 1962, **6**, 17–33.

Shankweiler, D. P. Effects of success and failure instructions on reaction time in brain-injured patients. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 546–549.

Silverman, J. Noxious cue sensitivity in schizophrenia. Paper read at American Psychological Association, Philadelphia, August 1963.

Slechta, Joan, Gwynn, W., & Peoples, C. Verbal conditioning of schizophrenics and normals in a situation resembling psychotherapy. *Journal of Consulting Psychology,* 1963, **27**, 223–227.

Smock, C. D., & Vancini, J. Dissipation rate of the effects of social censure in schizophrenics. *Psychological Reports,* 1962, **10**, 531–536.

Spohn, H. E., & Wolk, W. Effect of group problem solving experience upon social withdrawal in chronic schizophrenics. *Journal of Abnormal and Social Psychology,* 1963, **66**, 187–190.

Stotsky, B. Motivation and task complexity as factors in the psychomotor responses of schizophrenics. *Journal of Personality,* 1957, **25**, 327–343.

Sullivan, H. S. The language of schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. California Press, 1946. Pp. 4–16.

Tilton, J. R. The use of instrumental motor and verbal learning techniques in the treatment of chronic schizophrenics. *Dissertation Abstracts,* 1956, **16**, 1180–1181.

Topping, Gillian, & O'Connor, N. The response of chronic schizophrenics to incentives. *British Journal of Medical Psychology,* 1960, **33**, 211–214.

Turbiner, M. Choice discrimination in schizophrenic and normal subjects for positive, negative, and neutral affective stimuli. *Journal of Consulting Psychology,* 1961, **25**, 92.

Waters, T. J. Censure reinforcement, cue conditions and the acute-chronic schizophrenia distinction. Unpublished doctoral dissertation, University of Missouri, 1962.

Webb, W. W. Conceptual ability of schizophrenics as a function of threat of failure. *Journal of Abnormal and Social Psychology,* 1955, **50**, 221–224.

Weckowicz, T. E., & Blewitt, D. B. Size constancy and abstract thinking in schizophrenic patients. *Journal of Mental Science,* 1959, **105**, 909–934.

Werner, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.

White, Mary A. A study of schizophrenic language. *Journal of Abnormal and Social Psychology,* 1949, **44**, 61–74.

Whiteman, M. The performance of schizophrenics on social concepts. *Journal of Abnormal and Social Psychology,* 1954, **49**, 266–271.

Williams, E. B. Deductive reasoning in schizophrenia. Unpublished doctoral dissertation, Columbia University, 1962.

Winer, H. R. Incidental learning in schizophrenics. *Dissertation Abstracts,* 1954, **14**, 1002–1003.

Wing, J. K., & Freudenberg, R. K. The responses of severely ill chronic schizophrenics to social stimulation. *American Journal of Psychiatry,* 1961, **118**, 311–322.

Wittman, Phyllis. An evaluation of opposed theories concerning the etiology of so-called "dementia" in dementia praecox. *American Journal of Psychiatry,* 1937, **93**, 1363–1372.

Zahn, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. *Journal of Abnormal and Social Psychology,* 1959, **58**, 39–47.

Zaslow, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *Journal of Consulting Psychology,* 1950, **14**, 335–339.

Zigler, E., & Phillips, L. Social competence and the process-reactive distinction in psychopathology. *Journal of Abnormal and Social Psychology,* 1962, **65**, 215–222.

(Received March 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 25–31

THE EFFECTS OF SUCCESS AND FAILURE ON THE DISCRIMINATION LEARNING OF NORMAL AND RETARDED CHILDREN[1]

EARL C. BUTTERFIELD AND EDWARD ZIGLER
Yale University

30 normal and 30 retarded children learned a 3-choice size discrimination following exposure to experimental tasks under conditions of success, failure, or neutrality. Both normal and retarded children learned more quickly following success and failure than under the neutral condition. In a 2nd study, 32 normal and 32 retarded children learned the same criterion task under success or failure conditions. Here, both success and failure again facilitated the performance of both normal and retarded Ss. Also, social responsivity interacted with intelligence and experimental conditions in influencing the children's learning.

Considerable attention has been given to the effects of success and failure on the subsequent performance of normal and retarded children (Cromwell, 1963; Gardner, 1958; Heber, 1957; Kass & Stevenson, 1961; Steigman & Stevenson, 1960; Stevenson & Zigler, 1958). Two general themes can be found in the recent research efforts. The first is that success and failure experiences influence the general drive level of the child. Employing the Taylor-Spence (Taylor, 1956) drive formulation, the adherents of this view assert that changes in drive level, induced by both success and failure, either facilitate or attenuate performance depending upon the complexity of the task (Stevenson, 1960). The second theme is that the life histories of normal children are characterized primarily by success experiences, while the histories of retarded children are characterized primarily by failure experiences (Cromwell, 1963). The proponents of this view maintain that these histories result in differing expectancies of success which in turn underlie a wide variety of behavioral differences between normal and retarded children. This view further suggests that, as a result of their differing histories, normal and retarded children react differently to success and failure experiences.

These two themes are not necessarily antithetical. However, they do differ in that the first focuses on a general drive state while the second holds that any comprehensive understanding of the effects of success and failure experiences demands that the child's previous history of success and failure be taken into account. While these views are fairly straightforward, the findings emanating from their test have been far from consistent. In comparisons of normal subjects, one study reports that a success condition facilitates subsequent performance more than a failure or control condition with no differences between the latter two conditions (Kass & Stevenson, 1961). Another study found that failure attenuated performance beneath the level found in the success and control conditions and found no difference between the latter conditions (Stevenson & Pirojnikoff, 1958). Employing only success and failure conditions, investigators have found that a success condition both facilitated performance (Steigman & Stevenson, 1960) and attenuated performance (Stevenson & Zigler, 1958) as compared to a failure condition.

Studies involving performance of retarded children have also been characterized by inconsistency. In addition, the findings have been incongruent with those reported above. One investigator found that retardates performed better following success and poorer following failure as compared to a no manipulation control group (Gardner, 1957). Another investigator found that the performance of normals and retardates was equally enhanced following failure; and that while success enhanced the performance of both groups, the performance of retardates was enhanced more than that of normals (Heber, 1957). Other investigators found that success enhanced the performance of normals more than that of retardates (Kass & Stevenson, 1961). Another study also found that failure had a general enhancing effect for both normals and retardates but that failure enhanced the performance of normals more than that of retardates (Gardner, 1958).

---

[1] This research was supported by Research Grant MH-06809 from the National Institute of Mental Health, United States Public Health Service. The authors wish to thank the staff of the Southbury Training School and the New Haven Public School System for their cooperation.

It is extremely difficult to reconcile the inconsistencies in this body of work, especially since the studies differ in: (*a*) the method of inducing success and failure experiences, i.e., experimentally manipulating the amount of objective success or failure experienced and/or having the experimenter tell the child that he is doing well or poorly; (*b*) the degree of success and failure involved in the experimental manipulations; (*c*) the degree of similarity between the task upon which the subject received success or failure and the criterion measure; and (*d*) the criterion measures themselves. Despite these methodological differences, it appears extremely unlikely that a single general process, e.g., drive level of the subject, underlies the disparate findings reported above. This very disparity indicates that certain more idiographic variables influence the reactions of children to success and failure experiences.

The primary goal of the present study was to investigate whether retardates, as compared to normal children, do indeed react differently following failure experiences. Normal and retarded children were given success, failure, or no experience on two games and then tested on a partially reinforced discrimination learning task. All subjects received the same amount of objective success during the success-failure manipulations, the only difference between success and failure groups being the experimenter's comments concerning subjects' performance. The decision to employ a partially reinforced learning task was dictated by three factors: (*a*) Gardner (1958) employed as his criterion measure improvement on a nonlearning task which raises the question of whether his findings can be generalized to conventional learning situations; (*b*) a partially reinforced discrimination learning task has been found to be sensitive to the type of variable being investigated in the present study (Stevenson & Zigler, 1958); and finally, (*c*) the belief that a situation in which the child is not always reinforced for a correct response is characteristic of his real life experiences.

STUDY I

Method

Subjects. The subjects for this study were three groups of 10 institutionalized mentally retarded children and three groups of 10 normal children. The six groups were matched on MA, and each contained five boys and five girls. The mean CA and MA of each group are presented in Table 1. The retarded subjects were all of the familial type, and no individuals with gross sensory or motor disturbances were used. The Peabody Picture Vocabulary Test, Form B (Dunn, 1959), was used to obtain all MA ratings.

Procedure. The subjects were tested individually. As the experimenter brought the subject into the room (an office for the retarded subjects and an unused stage dressing room for the normals) the experimenter pointed to a variety of inexpensive toys which were spread on a table and said:

TABLE 1

CA AND MA IN MONTHS AND TRIALS TO CRITERION FOR EACH GROUP

Groups	*N*	CA		MA		Trials to criterion	
		M	*SD*	*M*	*SD*	*M*	*SD*
Retarded-control	10	162	29.9	88	14.0	39.1	13.5
Retarded-success	10	166	21.4	87	12.6	31.2	17.1
Retarded-failure	10	164	22.3	87	13.0	30.0	18.3
Normal-control	10	91	10.4	89	12.0	42.2	9.4
Normal-success	10	88	11.6	88	13.2	28.1	18.0
Normal-failure	10	90	15.2	89	14.1	27.5	22.0

> See, I have some toys and things here. I'm going to give you a chance to win one of these today. Which one do you like best? Which one would you most like to have?

After the subject made his choice, the experimenter said:

> Fine. Now we're going to play some games. By playing the games well, you can win marbles. If you win enough marbles, then you can have this (indicating the toy of the subject's choice).

Then the experimenter introduced the first experimental game. (All subjects played Pick-a-Card first and Which-School second.) Following the first game, the experimenter said: "Well, that's all of that game. Now here's the next game." The experimenter then presented the materials and gave the instructions for the second game. At the conclusion of the second experimental game, the experimenter said: "Well, that's all of that game. Now let's play the last game." The learning task was then presented and the subject played it until he reached a criterion of 10 consecutive correct responses or until he had been presented with 50 different arrangements of the stimuli. Following this, the subject was instructed to count the marbles he had won and was told that he had five more than he needed to win the prize. Each subject was then given his prize, praised liberally, and returned to his classroom or cottage.

The subjects were assigned to the three reinforcement conditions so that the retarded and normal groups would have comparable MAs, and each group would include an equal number of boys and girls. In the success condition, the experimenter made four positive statements to the subject at predetermined points during each of the experimental games. These statements always occurred on a response for which the subject was given a marble. The positive statements were: "That's good," "You really know how to play this game," "That's fine," "You really know how to play (name of game)." In the failure condition, the experimenter made four negative statements at predetermined points during both of the experimental games. These statements always occurred following a response for which the subject was not given a marble. These statements were: "That's not good," "You don't play this game very well," "That's bad," "You don't play (name of game) very well." In the control condition, the experimenter made no statements to the subject during the experimental games. All groups received the same amount of objective success, i.e., 16 marbles in each experimental game. Nothing was said to any subject during the criterion task except to repeat instructions about picking up only one stimulus.

Experimental games. The two experimental games used in this study were patterned after those used by Stevenson and Zigler (1958). Both games were quite simple, and performance depended primarily upon compliance with the experimenter's instructions.

Pick-a-Card. In this game the subject's task was to turn each of 24 1½ inch × 3 inch cards over and hand them to the experimenter one at a time. Sixteen of the cards had a red spot on one side and 8 of the cards had a blue spot. The deck of 24 cards was spread, plain side up, before the subject, and he was instructed:

> Here's the first game. It's called Pick-a-Card. Here are the cards. Some have blue spots on one side and some have red spots on one side. Your job is to pick up cards whose spots are the same colors as the colors I have written down here on this piece of paper. Sometimes a blue one is right and sometimes a red one is right. I'll spread the cards all over the table, like this. When it's time for you to begin, you'll guess by turning the cards over one at a time and handing them to me. Every time you hand me a right colored card, I'll give you a marble. If you have enough marbles after we have played three games, you can have this [indicating the object of child's preference]. You can keep the marbles in this marble board, and we'll count them after we have played three games to see if you have won enough marbles. All right, now you may begin picking up the cards, turning them over, and handing them to me one at a time. When you give me a right card, I'll give you a marble to put in your board.

In the course of turning over the 24 cards, all subjects were given 16 marbles in the same randomly predetermined schedule.

Which-School. In this game the subject's task was to place each of 24 3 × 5 inch glossy pictures of children into one of two boxes constructed to resemble buildings. One "building" was painted green and the other was painted pink. The boxes were placed about 6 inches apart and 12 inches in front of the subject with their open sides facing the subject. The experimenter said:

> Now, here's the next game. It's called Which-School. Here are the two schools. One is pink and one is green. Here are some pictures of children. Each of these children goes to one of these schools. Your job is to guess which school each child goes to. You guess by taking the pictures one at a time and putting them in the school you think that child goes to. When you put the picture in the right school, I'll give you a marble. Here are the pictures. You put them in the schools one at a time.

A random schedule was employed in rewarding the children with a marble on 16 of the 24 responses.

Criterion task. Although the criterion task was presented to the subjects as just another game in which they could win marbles, it differed from the experimental games in that securing marbles was dependent upon learning a correct response. The criterion measure was a three-choice, size-discrimination learning problem. The subject's task was to find marbles by picking up one of three different size blocks each time they were presented to him. The stimuli were 3 inch, 4½ inch, and 6 inch square blocks cut from ½ inch plyboard and painted flat black. The stimuli were presented to the subject in a predetermined random order. They were placed on a board containing marble holes which was pushed towards the subject from beneath a raised 24 × 36 inch inclined shield. This shield prevented the subject from seeing either the stimuli or the experimenter's movements when the board was pulled toward the experimenter. The subject was allowed to pick up and look under only one block on each presentation. The order of the blocks was changed only after those presentations upon which the subject chose the correct block. Following an incorrect choice, the experimenter went through all of the motions of changing the positions of the blocks but actually did not.

The stimulus to be rewarded was randomly determined for each subject, with the restriction that each of the three stimuli was correct for approximately one-third of the subjects in the experimental group. A subject found a marble under the correct stimulus two out of every three times he chose that stimulus. The same randomly predetermined arrangements of stimuli and reinforcement schedule were used for every subject.

The instructions to the subjects were:

> Now we're going to play the last game. It's called Find-the-Marble. By playing well, you can win lots of marbles. If you win enough marbles, remember, you can have this [name the child's preferred toy]. Now, let me show you how to play this game. See these three difference size blocks. One is big, one is little, and one is middle size. I'm going to hide a marble under one of these blocks. I will always hide it under the same block. It will never be under the other two blocks. But, every time I hide the marble, I'll mix the blocks up, like this. So, each time I push the blocks out, you'll have to look at all three of them to find the one where I have hidden the marble. Remember, I'll always put the marble under the same block. It will never be under the other two. But, every time I hide the marble, I'll mix the blocks up so that you will have to look at all three blocks to find the one where I have hidden the marble. Every time I push the blocks out, you can look underneath one of them.

Results and Discussion

The mean number of trials to criterion for the normal and retarded groups under each of the three conditions are presented in Table 1. As with the conventional correction procedure, a trial was defined by a correct response. Analyses of data involving the absolute number of stimulus presentations resulted in essentially the same findings as reported below. (Preliminary analyses revealed no significant sex effects, and for the purposes of clarity the sex variable is not included in the following analyses.) A 2 × 3 factorial analysis of variance revealed a significant main effect for experimental conditions ($F_{2/54} = 3.2$, $p < .05$), no significant effect for type of subject ($F < 1$) and no significant interaction effect ($F < 1$). Further analysis disclosed that the experimental conditions effect was due to a smaller mean trials to criterion in both the failure ($t = 2.58$, $p < .02$) and success ($t = 2.60$, $p < .02$) conditions than in the control condition. No significant difference was found between the failure and success conditions ($t < 1.0$).

The results clearly indicate that prior experiences of both success and failure facilitate performance of both normal and retarded children. The findings offer no support for the view that success and failure differentially affect children's performance, nor for the view that normals and retardates react differently to success and failure. These findings appear to call into question the anxiety interpretation advanced in earlier studies, unless one assumes that the success and failure manipulations both induce anxiety and that the criterion task employed is of such a simple sort that performance would be facilitated by anxiety. Neither of these assumptions appears plausible.

The finding that both success and failure facilitated performance can be explained by the following position: the child typically satisfies a variety of motives in the experimental situation, e.g., curiosity, having fun with the "game," interacting with an attending adult, avoiding unpleasant classroom activities, obtaining a tangible reward, and successfully solving the problem. What must be emphasized is that being successful, as defined by performing the correct response, is just one of many motives and may, in fact, compete with others. Given this complex motive state, either a success or failure manipulation causes the child to orient himself towards success or failure as defined by the experimenter with other motives receding in importance. The success-failure orientation, having become paramount, results in superior problem solving, the criterion of success which the child notes is being employed by the experimenter.

This general view concerning the effects of success-failure manipulations cannot subsume the findings that success and failure experiences not only have different effects, but have differentially affected the performance of normal and retarded children. Apparently, some more idiographic factor(s) must be considered before we can predict how any particular normal or retarded child will react to success or failure manipulations. One such factor is suggested in the general position advanced above. While inducing a success-failure orientation raises the success motive in the child's motive hierarchy, just how dominant this motive becomes will probably depend upon its relative position in the hierarchy prior to the success-failure manipulations. Stated simply, the greater the child's intrinsic desire to win the approval of the experimenter, the greater will be the facilitating effect of the success-failure manipulations. The test of such a view requires an independent measure of the child's motivation for adult approval. While no such measure was available for subjects in Study I, the authors had access to a pool of subjects for whom there were measures of the effectiveness of adult approval as a reinforcer. In Study II, retarded and normal subjects who were either responsive or unresponsive to adult approval were given either success or failure manipulations prior to learning the same partially reinforced discrimination problem employed in Study I.

STUDY II

Method

Subjects. Four groups of eight institutionalized, mentally retarded children and four groups of eight normal children were employed. The retarded subjects were all of the familial type, and no individuals with gross sensory or motor disturbances were used. The groups were matched on MA (see Table 2) and each group contained an equal number of boys and girls. The Peabody Picture Vocabulary Test, Form B (Dunn, 1959) was used to obtain MA ratings. The CA and MA of the subjects in Study II was somewhat higher than those of Study I (see Tables 1 and 2). Earlier research (Zigler, 1958) indicates that differences of this magnitude within this age range do not appreciably affect scores on the persistence task.

A normal and a retarded group were selected from children who were highly responsive to social reinforcement, and a normal and a retarded group were selected from children who were relatively unresponsive to social reinforcement. Responsivity to social reinforcement was defined by the length of time the child played a monotonous two-part task under conditions of positive social reinforcement. The task employed was a two-hole marble game which has frequently been used to assess the effectiveness of social reinforcement (see Zigler, 1961 for a full description of this measure). The responsive and unresponsive subjects were randomly selected from the upper and lower quartiles of populations of approximately 200 normal and 200 retarded subjects who had been run on the social reinforcement task during the preceding year in the course of other research projects. It was assumed that these scores were still valid indicators of the subjects' responsivity to social reinforcements. Half of each Responsivity × Intelligence grouping was given success manipulation and half was given failure manipulation so that a 2 × 2 × 2 factorial design was formed with intel-

TABLE 2

CA AND MA IN MONTHS AND TRIALS TO CRITERION FOR EACH OF THE GROUPS IN STUDY II

Groups	*N*	CA		MA		Trials to criterion	
		M	*SD*	*M*	*SD*	*M*	*SD*
Retarded							
Failure condition							
High responsivity	8	193	48.6	106	36.4	39.0	17.9
Low responsivity	8	193	39.2	111	23.0	18.1	17.6
Success condition							
High responsivity	8	211	35.4	113	26.5	20.3	17.9
Low responsivity	8	210	38.3	118	20.0	32.2	30.9
Normal							
Failure condition							
High responsivity	8	98	16.3	97	20.0	14.3	11.1
Low responsivity	8	96	17.8	95	22.6	20.8	17.2
Success condition							
High responsivity	8	97	18.1	97	26.3	25.8	29.1
Low responsivity	8	94	16.6	94	21.2	18.0	8.1

ligence, responsivity, and experimental condition as the dimensions.

Procedure. Exactly the same procedures and criterion scores were used as in Study I, except that no prizes were offered to the subjects and the neutral condition was omitted. It became apparent during Study I that the receipt of marbles alone was highly motivating.

Results and Discussion

Ignoring the responsivity variable, the criterion scores obtained by the success and failure groups in Study II were quite comparable to those obtained in Study I (see Tables 1 and 2). Simple randomized analyses of variance showed no significant differences between normal and retarded subjects in Study I and Study II under either success ($F < 1$) or failure ($F_{3/48} = 2.0$, $p > .10$) conditions.

Employing the control group of Study I ($\bar{X} = 40.6$) and again ignoring the responsivity variable, it was found that the success manipulations facilitated the performance of both normal ($\bar{X} = 22.5$, $t = 4.30$, $p < .01$) and retarded ($\bar{X} = 26.3$, $t = 2.26$, $p < .05$) subjects. The failure manipulation also facilitated the performance of both normal ($\bar{X} = 17.5$, $t = 4.27$, $p < .01$) and retarded ($\bar{X} = 28.5$, $t = 2.66$, $p < .01$) subjects. These results are in agreement with the findings of Study I that both success and failure conditions facilitate the performance of both average and retarded subjects. This general facilitating effect should be kept in mind when assessing the role of the social responsivity variable.

The effects of the social responsivity variable were evaluated in the 2 × 2 × 2 factorial design described earlier (see Table 3). This analysis revealed a significant Intelligence × Responsivity × Condition interaction (see Figure 1). None of the other effects were significant.

TABLE 3

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION

Source	*df*	*MS*	*F*
Between			
Type of subject (A)	1	877.6	2.14
Success-failure (B)	1	26.2	<1
Responsivity (C)	1	87.9	<1
A × B	1	199.6	<1
A × C	1	74.4	<1
B × C	1	489.6	1.19
A × B × C	1	2,103.3	5.13*
Within	56	410.0	
Total	63		

* $p < .05$.

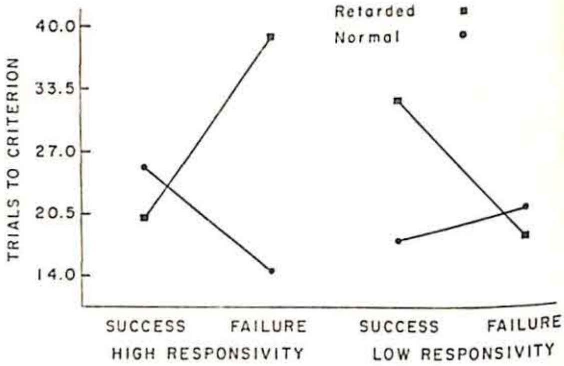


FIG. 1. Mean trials to criterion for high and low responsive groups following success and failure manipulations.

As can be seen in Figure 1, the nature of this significant triple interaction is quite complex. A variety of post hoc explanations could be advanced to explain this particular pattern of findings. The one clear conclusion that can be drawn is that for both normals and retardates, the child's responsivity to social reinforcement influences the effects of success and failure manipulations. Failure to control for such idiographic variables as responsivity to social reinforcement may account for the diversity of findings in the success-failure literature.

Some more tentative conclusions may be derived from the findings shown in Figure 1. A striking difference was found between retarded and normal responsive groups following failure experiences. No such difference was found between the retarded and normal nonresponsive groups. This difference among responsive subjects lends some support to the view that retardates and normals react differently to failure experiences. However, contrary to certain earlier findings, the retardates' performance following failure was not impaired, but rather was not facilitated. It would appear that responsive normal children react to failure experiences with the increased effort that accompanies the hypothesized success-failure orientation. Why responsive retarded children show no such facilitation remains an open question. Two hypotheses can be advanced. The first is that failure experiences arouse competing motives in responsive retarded children which they do not arouse in responsive normals. For example, failure experiences may be so threatening, i.e., anxiety-producing to retardates, that it causes them to deny that a failure experience did in fact occur. One would not expect a child who denied the existence of a failure experience to react to it with the increased effort that accompanies a success-failure orientation.

A second, and perhaps more plausible, hypothesis has been advanced by Cromwell and his students (Cromwell, 1963). These investigators have suggested that while increased effort is a common response to failure, such a response has been reinforced more frequently in the life experiences of normal than in those of retarded children. This view generated the prediction that normals were more likely than retardates to enhance their performance following failure experiences, a prediction that was confirmed by Gardner (1958). Consistent with this view were the findings of Shallenberger and Zigler (1961) that failure experiences led to increased persistence by normals but not by retardates. The findings of the present study suggest that this argument may hold for socially responsive children rather than for normal and retarded children in general.

REFERENCES

CROMWELL, R. L. A social learning approach to mental retardation. In N. R. Ellis (Ed.), *Handbook of mental retardation.* New York: McGraw-Hill, 1963.

DUNN, L. M. *Peabody Picture Vocabulary Test.* Minneapolis: American Guidance Service, 1959.

GARDNER, W. I. Effects of interpolated success and failure on motor task performance in mental defectives. Paper read at Southeastern Psychological Association, Nashville, March 1957.

GARDNER, W. I. Reactions of intellectually normal and retarded boys after experimentally induced failure: A social learning theory interpretation. Unpublished doctoral dissertation, George Peabody College, 1958.

HEBER, R. F. Expectancy and expectancy changes in normal and mentally retarded boys. (Doctoral dissertation, George Peabody College) Ann Arbor, Mich.: University Microfilms, 1957, No. 24475.

KASS, N., & STEVENSON, H. W. The effect of pretraining reinforcement conditions on learning by normal and retarded children. *American Journal of Mental Deficiency,* 1961, **66**, 76–80.

SHALLENBERGER, PATRICIA, & ZIGLER, E. Rigidity, negative reaction tendencies, and cosatiation effects in normal and feebleminded children. *Journal of Abnormal and Social Psychology,* 1961, **63**, 20–26.

STEIGMAN, M. J., & STEVENSON, H. W. The effects of pretraining reinforcement schedules on children's learning. *Child Development,* 1960, **31**, 53–58.

STEVENSON, H. W. Learning of complex problems by normal and retarded subjects. *American Journal of Mental Deficiency,* 1960, **64**, 1021–1026.

STEVENSON, H. W., & PIROJNIKOFF, L. A. Discrimination learning as a function of pretraining reinforcement schedules. *Journal of Experimental Psychology,* 1958, **56**, 41–44.

STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *Journal of Experimental Psychology,* 1958, **56**, 185–192.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychological Bulletin,* 1956, **53**, 303–318.

ZIGLER, E. The effect of pre-institutional social deprivation on the performance of feebleminded children. Unpublished doctoral dissertation, University of Texas, 1958.

ZIGLER, E. Social deprivation and rigidity in the performance of feebleminded children. *Journal of Abnormal and Social Psychology,* 1961, **62**, 413–421.

(Received January 24, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 32–33

RELIABILITY OF CLINICAL JUDGMENTS AS A FUNCTION OF RANGE OF PATHOLOGY[1]

WILLIAM A. HUNT, MELVIN L. SCHWARTZ
Northwestern University
AND RONALD E. WALKER
Loyola University, Chicago

The relationship between the reliability of clinical judgment and the severity of the pathology being judged is investigated using 6 sets of data involving a total of 2055 stimuli and 41 judges. The stimuli were intelligence test items obtained from patients and were judged by experienced clinicians using a 7-point scale to estimate the amount of confusion in thinking exhibited in the response. The hypothesis, that stimuli at the extremes (least and most) of the scale would be judged more reliably than those in the middle, was confirmed in all 6 instances.

In an informational analysis of clinical judgment, Miller and Bieri (1963) note "the rather good agreement by the subjects in regard to the two extreme cases in each stimulus domain," and note this agreement with psychophysical research where "extreme stimuli are judged more reliably than are middle-range stimuli." They further suggest "it would be of interest to pursue this notion in further research by systematically varying the range of pathology from which the stimulus materials are to be drawn."

During an extensive study of the factors influencing clinical judgment (Hunt & Jones, 1962), the authors have accumulated extensive data involving judgments on a seven-point scale of the amount of confusion in patients' thinking as revealed in their individual test responses. These data are amenable to analysis in terms of the standard deviation of the judgments for any scale point. If the relationship noted by Miller and Bieri is a general one, then the standard deviations would be expected to be less at the extremes of the scale and larger in the middle ranges.

Table 1 reports the standard deviation of the judgments for each scale interval from 1.00–1.99, 2.00–2.99, etc., to 6.00–6.99, 6 intervals in all. Each entry shows the mean standard deviation for all stimuli whose mean scale value fell within that interval. Each stimulus was judged by the same number of experienced clinicians (PhD plus 5 years of professional experience). The first set of data comes from our earliest standardization series (Hunt & Arnhoff, 1955) in which 50 vocabulary test responses were rated by 16 judges; the second set from an experiment (Hunt, Walker, & Jones, 1960) in which 5 judges scaled each of 383 vocabularly responses; the third set from the same experiment is based on 5 judges each of whom judged 141 comprehension test responses; the fourth set involves the judgments of 5 judges for 327 comprehension responses (Hunt & Walker, 1962); the fifth set is from the same experiment and involves 5 judges and 883 vocabulary responses; and the sixth set (Schwartz, Hunt, & Walker, 1963) includes 5 judges judging 271 vocabulary responses. Sets 1, 2, 3, and 6 involved schizophrenic subjects. Sets 4 and 5 were gathered on a mixed group of 30 cases from five diagnostic categories. In all 6 sets of data the hypothesized relationship holds with the smaller standard deviations appearing at the extremes of the scale. It is noteworthy that in no set is there any reversal in the orderly progression of the values.

Obviously artifact is contributing to the results since truncation of the scale at the ends should reduce the standard deviation in the extreme categories. Careful inspection of the data, however, reveals that while this

---

[1] This study is part of a larger project being conducted at Northwestern University under the direction of William A. Hunt through contract Nonr-1228(18) with the Office of Naval Research.

TABLE 1

STANDARD DEVIATION OF JUDGMENTS AS A FUNCTION OF SCALE VALUE OF STIMULUS

Experiment	Mean scale values					
	1.00–1.99	2.00–2.99	3.00–3.99	4.00–4.99	5.00–5.99	6.00–6.99
1	.62	.94	1.14	1.16	1.06	.67
2	.32	1.29	1.67	1.66	1.55	.86
3	.33	1.11	1.25	1.54	1.32	.85
4	.54	1.01	1.19	1.06	1.03	.68
5	.48	1.09	1.23	1.20	1.04	.70
6	.46	.89	1.06	1.05	.92	.46

seems true at the low end of the scale, at the high end the judges have anchored the scale well outside the range of stimuli provided in the experiment. Thus the scale category "1" is commonly used and there are many stimuli with a mean value of 1 indicating unanimous agreement among all the judges in that particular experiment; but the scale value "7" is rarely used and no stimulus has ever achieved a mean value this high. This should alleviate the truncation although all the effects of such self-instructed anchoring are not completely understood (Campbell, Lewis, & Hunt, 1958). It should be noted that the stimuli, gathered from many schizophrenics in many hospitals, actually do offer a complete range of pathology.

Some of the stimulus series represented in Table 1 are "natural" samplings consisting of all the responses given by a selected group of patients. In such instances the distribution of scale values is not normal but j shaped with a piling up of values in the lower categories. Here one might suppose that the greater practice with Category 1 might reduce the variability in its use and give a gradual increase in standard deviations from 1 through 7. Obviously such is not the case. In other series, the number of stimuli is equated for each scale value. The same reliability function holds as with the natural samplings.

Considering the complexity of the task and the diversity of tests, judges, and diagnostic entities involved, the uniformity of the data is noteworthy, and confirms our belief in the orderly nature of the processes of clinical judgment (Hunt & Jones, 1962).

REFERENCES

CAMPBELL, D. T., LEWIS, NAN A., & HUNT, W. A. Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *Journal of Experimental Psychology,* 1958, **55**, 220–228.

HUNT, W. A., & ARNHOFF, F. W. Some standardized scales for disorganization in schizophenic thinking. *Journal of Consulting Psychology,* 1955, **19**, 171–174.

HUNT, W. A., & JONES, N. F. The experimental investigation of clinical judgment. In A. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962. Pp. 26–51.

HUNT, W. A., & WALKER, R. E. A comparison of global and specific clinical judgments across several diagnostic categories. *Journal of Clinical Psychology,* 1962, **18**, 188–194.

HUNT, W. A., WALKER, R. E., & JONES, N. F. The validity of clinical ratings for estimating severity of schizophrenia. *Journal of Clinical Psychology,* 1960, **16**, 391–393.

MILLER, H., & BIERI, J. An informational analysis of clinical judgment. *Journal of Abnormal and Social Psychology,* 1963, **67**, 317–325.

SCHWARTZ, M. L., HUNT, W. A., & WALKER, R. E. Clinical judgment of vocabulary responses in process and reactive schizophrenia. *Journal of Clinical Psychology,* 1963, **19**, 488–494.

(Received January 28, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 34–37

RATE OF BINOCULAR RIVALRY ALTERNATION IN PSYCHOTIC AND NONPSYCHOTIC PATIENTS

ROBERT FOX
Vanderbilt University

To determine the relationship between binocular rivalry rate and the presence of psychopathology, rivalry rate measures for 2 1-min. observation periods were taken from 3 matched groups of organic, schizophrenic, and nonorganic-nonpsychotic patients. The rate of the nonorganic-nonpsychotic group increased significantly over observation periods relative to the other groups, a result consistent with other data which suggests that rate provides some measure of cognitive efficiency. Differences in rate between organic and schizophrenic groups were not found, contrary to previous report. Procedural variables which could account for the discrepancy are discussed.

Binocular rivalry is a centrally controlled phenomenon closely related to attentional processes; consequently, some interest has been shown in the identification of variables which influence rate. Reports by George (1936), Bárány and Hallden (1947), Platz, Uhr, and Miller (1960), and Ruttiger (1963) indicate that depressants (e.g., alcohol) and tranquillizing drugs (e.g., meprobamate) reduce rate while stimulating drugs (e.g., caffeine) increase rate. Positive correlations between rate and intelligence test scores within a college population have been obtained by Crain (1961). Also, there are data which suggest an inverse relation between rate and degree of psychopathology. Eysenck, Granger, and Brengelmann (1957) found such an ordering of the mean rates for groups of normal, neurotic, and psychotic patients. Sappenfield and Ripke (1961) obtained striking differences in rate between organic, schizophrenic, and nonhospitalized control subjects. With reference to a combined median cutting score, they report that only 1 of the 30 organic patients exhibited a rate as high as the control subjects and only 4 organics reached the level of the schizophrenic distribution.

The purpose of this note is to report additional data relevant to the relationship between rate and psychopathology and, further, to call attention to the potential use of rivalry rate as an indicator of individual differences. Specifically, the present experiment compared the rivalry rate of three matched groups of organic, schizophrenic, and nonpsychotic-nonorganic patients.

METHOD

Subjects. The subjects were newly admitted patients of a large state mental hospital[1] who had not yet been assigned drug therapies. Three groups of 15 subjects each were assembled on the basis of diagnosis and upon compliance with matching variables. These variables were age, sex, level of education (last grade completed), and verbal IQ (defined as scale score on the vocabulary subtest of the WAIS).

Subjects were also tested for visual functioning on five tests of the Modified Orthorater (Bausch & Lomb) and rejected if corrected far-point acuity in the weakest eye was less than 3 or if phoria of greater than 11 or less than 2 was present. This screening resulted in a posteriori matching of groups on the five Orthorater tests employed: far-point acuity for binocular, and left- and right-monocular conditions, far-point lateral phoria, and vertical phoria. Two to four potential subjects from each of the groups were rejected because of visual insufficiency.

The diagnostic categories represented within each of the three groups were as follows: for the *organic* group, chronic brain syndrome (CBS) with alcoholism (8), CBS with trauma (3), CBS with CNS syphilis (1), CBS with neurotic depression (1), and CBS with behavioral reaction (2); for the *schizophrenic* group, schizophrenic reaction, chronic undifferentiated (6), schizophrenic reaction, paranoid type (7), and schizophrenic reaction, schizo-affective type (2); for the *nonpsychotic-nonorganic* group, personality trait disturbance, passive aggressive (4), personality trait disturbance, emotionally unstable (3), sociopathic personality disturbance, dysocial reaction (3), sociopathic personality disturbance with alcoholism (3), and psychoneurotic anxiety reaction (2). Diagnosis was based upon clinical assessments independently made by the psychological and psychiatric staffs. The subjects were selected only if agreement among diagnosticians was present.

---

[1] Appreciation is expressed towards the members of the Psychology Department of Longview State Hospital, Cincinnati, who assisted in the procurement and running of subjects.

Means and standard deviations of the matching variables for each group are given below. The difference in age between the organics and nonpsychotic-nonorganic groups is significant ($p < .05$); other differences are not significant.

Organic: 12 males, 3 females; age $M = 42.1$, $\sigma = 10.1$; education level $M = 7.5$, $\sigma = 2.8$; vocabulary score $M = 7.7$, $\sigma = 1.8$; acuity: binocular $M = 8.8$, right eye $M = 8.1$, left eye $M = 8.1$; phoria: lateral $M = 6.9$, vertical $M = 5.7$.

Schizophrenic: 10 males, 5 females; age $M = 40.3$, $\sigma = 9.2$; education level $M = 9.4$, $\sigma = 2.1$; vocabulary score $M = 8.9$, $\sigma = 2.0$; acuity: binocular $M = 9.5$, right eye $M = 7.9$, left eye $M = 7.7$; phoria: lateral $M = 7.8$, vertical $M = 5.0$.

Nonpsychotic-nonorganic: 12 males, 3 females; age $M = 37.1$, $\sigma = 9.4$; education level $M = 8.6$, $\sigma = 2.1$; vocabulary score $M = 8.7$, $\sigma = 1.9$; acuity: binocular $M = 8.4$, right eye $M = 8.2$, left eye $M = 7.4$; phoria: lateral $M = 5.9$, vertical $M = 5.9$.

Apparatus. The apparatus consisted of the Modified Orthorater, which is essentially a Brewster-type stereoscope. To induce rivalry a 2.5×7.5 inch translucent Plexiglas slide was fitted for the far vision holder of the Orthorater. Red and green theatrical filters (Edmund Scientific) were placed on the slide to fill the respective left and right eye fields. Each field was 2×2.5 inches; horizontal visual angle for each field was 5 degrees 30 minutes of arc. Without the colored filters the luminance level of the left field was 51.2 millilamberts, the right field was 42.3 millilamberts. To facilitate rivalry diagonal parallel lines $\frac{1}{4}$ inch apart were ruled on the filters and oriented in opposing directions.

Design and procedure. Each subject was first administered the five Orthorater tests of visual capacity. If the subject met the visual function criteria the rivalry slide was introduced as an apparent continuation of the preceding tests. The following instructions were read aloud.

> Now look at this slide. On one side there is a green square with slanting lines (monocular view). On the other side there is a red square with lines slanting the opposite way (monocular view). Now when you look at them (binocular view) what color do you see? (Wait for spontaneous comment.) Now I want you to tell me if the color of the square changes—for example, from red to green, or green to red—as you stare at the slide. If the color changes tell me by saying red or green, as the case may be. That is, say the name of the main color that you see whenever the color changes. Start telling me when I say "go," and keep it up until I say "stop."

All subjects appeared to understand the task and tended spontaneously to call out color names in serial order; e.g., "red, red, red, green, green, red. . . ." A few subjects used idiosyncratic terms to indicate changes; e.g., "stop" and "go" for red and green.

The basic datum was the number of changes reported by the subjects during two 1-minute observation periods. For the first period subjects actually observed for 75 seconds; the first 15 seconds, regarded as practice, was not recorded. The subject rested for 15 seconds with his head away from the eyepiece before beginning the second period. The experimenter timed the intervals with a stopwatch and recorded the reports of the subject. The rate score consisted of the number of transitions between states (i.e., red-to-green and green-to-red) during each of the 1-minute periods. This method of signaling and scoring changes is similar to that used in a number of previous studies, although it should be noted that it possesses several limitations. The method does not take into account indeterminate states in which both fields are partially suppressed, under the assumption that red and green responses are equally probable during the indeterminate period. In addition, no direct measure of the relative dominance of each field is provided. It is also necessary to assume that the use of verbal report as an indicator response does not involve special biases. These considerations suggest that more adequate rate measures would be provided by response keys or by the inhibition of optokinetic-nystagmus technique developed by Enoksson (1963).

Results and Discussion

In Table 1 the mean rates and standard deviations for the three groups for the first and second minutes are given.

The results of an analysis of variance of these data are given in Table 2. Variances are not heterogeneous, as assessed by Hart-

TABLE 1

Binocular Rivalry Rate Means and Standard Deviations for Organic, Schizophrenic, and Nonpsychotic–Nonorganic Patients

Group	First minute			Second minute		Combined	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Organic	15	16.7	8.54	18.3	8.69	34.9	16.51
Schizophrenic	15	16.4	7.19	18.7	7.60	35.1	13.75
Nonpsychotic-nonorganic	15	19.7	7.49	26.3	9.84	46.1	16.67

TABLE 2

ANALYSIS OF VARIANCE OF RIVALRY RATE SCORES

Source	*SS*	*df*	*MS*	*F*	*p*
Between subjects	6,174.822	44			
Groups	612.422	2	306.211	2.31	<.25
Error (a)	5,562.400	42	132.438		
Within subjects	985.500	45			
Trials	273.878	1	273.878	19.14	<.005
Groups × Trials	110.555	2	55.278	3.86	<.05
Error (b)	601.067	42	14.311		
Total	7,160.322	89			

ley's *F max* test. The significant interaction and trials effects are due to the large increase in rate from the first to the second minute exhibited by the nonpsychotic-nonorganic group. The results of a multiple-comparison analysis (Newman-Keuls, $p = .05$) indicate that between- and within-group differences for the organic and schizophrenic groups are not significant. The first-minute rate for the nonpsychotic-nonorganic group is not significantly greater than the second-minute rate for both the organic and schizophrenic groups. Product-moment correlations between the first and second minutes are as follows: organic group, $r = .84$; schizophrenic group, $r = .75$; nonpsychotic group, $r = .85$.

The higher rivalry rate exhibited by the nonorganic-nonpsychotic group is consistent with the suggestion of an inverse relation between rate and degree of psychopathology. However, the increase in rate in the present experiment became manifest only during the second observation period; it is difficult to determine from previous reports if such a temporal interaction has been obtained before, since the data previously reported have been averaged over trials.

The highly similar performance of the schizophrenic and organic groups is not congruent with the results obtained by Sappenfield and Ripke (1961). There are a number of differences in the procedure and apparatus employed in the two experiments and certain of these differences may be responsible for the discrepancy. The existing data on rivalry are insufficient to permit ready identification of the critical variables, although there is some basis for considering instructions to the subject as a probable factor. In the Sappenfield and Ripke experiment the instructions simply stated, "Tell me when the card you are viewing changes." Obviously, these instructions are less structured than those used in the present experiment. Research on the perception of the spiral aftereffect by organic patients indicates that such subjects tend to be reluctant to report the aftereffect, but will report it as frequently as other subjects when facilitating instructions are administered (London & Bryan, 1960; Mayer & Coons, 1960). The perception of both the aftereffect of movement and binocular rivalry involve novel and labile phenomenal changes, and it would not be surprising if some reluctance to report rivalry changes was present on the part of organic patients. In this regard it has been shown that manipulation of instructions can alter rivalry rate among normal subjects (Meredith & Meredith, 1962). This speculative interpretation is only one of several alternatives which can be entertained in view of the limited range of variables so far examined in rivalry experiments.

It should also be noted that even though differences in rate may be related to various diagnostic classifications, the interpretation of such rate variations is not unequivocal. While there are reasonable grounds for assuming that the rivalry mechanism is sensitive to organic conditions, information on the operation of the mechanism is not yet sufficient to provide a fully adequate theoretical structure. Further, the methods typically employed in measuring rate (including this experiment) have not minimized the potentially confounding contribution of response variables. For example, it is possible that variables which have been found to alter rate act upon the subject's criterion for reporting a change rather than upon the rivalry-switching mechanism itself. Until the influence of such variables is understood more completely, some caution appears to be required in the interpretation of rivalry rate differences.

REFERENCES

BÁRÁNY, E. H., & HALLDEN, U. The influence of some central nervous depressants on the reciprocal inhibition between the two retinae as manifested in retinal rivalry. *Acta Physiologica Scandinavica*, 1947, **14**, 296–316.

CRAIN, K. Binocular rivalry: Its relation to intelligence, and general theory of its nature and physiological correlates. *Journal of General Psychology,* 1961, **64**, 259–283.

ENOKSSON, P. Binocular rivalry and monocular dominance studied with optokinetic nystagmus. *Acta Ophthalmologica,* 1963, **41**, 544–562.

EYSENCK, H. J., GRANGER, G. W., & BRENGELMANN, J. C. *Perceptual processes and mental illness.* New York: Basic Books, 1957.

GEORGE, R. W. The significance of the fluctuations experienced in observing ambiguous figures and in binocular rivalry. *Journal of General Psychology,* 1936, **15**, 39–61.

LONDON, P., & BRYAN, J. M. Theory and research on clinical use of Archimedes spiral. *Journal of General Psychology,* 1960, **62**, 113–125.

MAYER, E., & COONS, W. H. Motivation and the spiral aftereffect with schizophrenic and brain-damaged patients. *Canadian Journal of Psychology,* 1960, **14**, 269–274.

MEREDITH, G. M., & MEREDITH, C. G. W. Effect of instructional conditions on rate of binocular rivalry. *Perceptual and Motor Skills,* 1962, **15**, 655–664.

PLATZ, A., UHR, L., & MILLER, J. G. A pilot experiment on the effects of meprobamate on stereoscopic retinal rivalry of complementary colors. *Perceptual and Motor Skills,* 1960, **10**, 230.

RUTTIGER, K. F. Individual differences in reaction to meprobamate: A study in visual perception. *Journal of Abnormal and Social Psychology,* 1963, **67**, 37–43.

SAPPENFIELD, B. R., & RIPKE, R. J. Validities of three visual tests for differentiating organics from schizophrenics and normals. *Journal of Clinical Psychology,* 1961, **17**, 276–278.

(Received January 28, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 38–46

INFLUENCE OF INSTRUCTIONS ON HYPNOTIC AND NONHYPNOTIC PERFORMANCE[1]

ROBERT SLOTNICK AND PERRY LONDON
Stanford University *University of Southern California*

This study examined the performance of high- and low-susceptible Ss under hypnotic and waking conditions when exhortative, analgesic, and other instructions were separately controlled and not confounded. 50 female Ss, assigned to 1 of 3 experimental or 2 control groups, performed 2 maximum effort tasks (Hand Dynamometer and Weight Endurance) and 1 skilled task (Tremor) in a Base Rate and a Special Instruction session. Except for 1 control group that always performed awake, all Ss performed both hypnotized and awake in both sessions. Special instructions facilitated performance in the hypnotic but not the waking state, and exhortative were more effective than analgesic instructions.

A number of experiments have studied the effects of hypnosis on simple performance tasks, but contradictory results among several of them have made it difficult to draw many definite conclusions. In one of the earliest studies, according to Hull's report (1933, pp. 226–227), Reiger (1884) found substantial improvement in endurance, measured by ability to hold the arm out horizontally while hypnotized. Nicholson (1920) found that hypnotized subjects far surpassed waking subjects in work capacity as measured by the ergograph. In the same vein, Hadfield (1924) reports increased grip-strength scores on a hand dynamometer during the hypnotic state. Williams (1930), however, contradicted Reiger's results in a better controlled study; he also modified Nicholson's claims, but conservatively supported the general trend. Young (1925, 1926), on the other hand, found no differences in hand dynamometer scores between hypnotic and waking states over several hundred measurements.

More recently, Eysenck (1941) found consistent improvement during hypnosis on physical performance tasks. Roush (1951) did a well controlled study; in Phase I, she found improvement during hypnosis on an arm dynamometer, but none either on a hand dynamometer or on an endurance test. In Phase II, she repeated the same tests but added instructions stating that subjects would feel no pain and would improve their performances. In this phase, performance improved significantly on all three tests. She attributes this improvement to the "no pain" instructions.

London and Fuhrer (1961) and Rosenhan and London (1963a), in a different design than the foregoing, used both high- and low-susceptible subject groups, and found performance on strength and endurance tasks better among low-susceptible subjects; they also found that both groups gave better performances in the waking than the hypnotic state. London and Fuhrer also added a special instruction condition which made it possible to differentiate between hypnosis proper and exhortation instructions. In that experiment, hypnosis alone tended to improve performance of high susceptibles only, exhortation improved everyone's performance, and the interaction between them was effective for one task (endurance) but not for another (strength).

In most previous experiments, exhortation, analgesia, and hypnotic instructions have all been confounded with each other. The present study examined task performance of high- and low-susceptible subjects under hypnotic and waking conditions when exhortation and pain instructions were separately controlled.

---

[1] This study was conducted in the Laboratory of Human Development, Stanford University, Ernest R. Hilgard, Director, and was supported in part by United States Public Health Service Grants MH 03859 (E. R. Hilgard, principal investigator) and MH 08598 (Perry London, principal investigator). The authors are indebted to E. R. Hilgard for aid in the design of the experiment and to Byron Klorfine for participation in the collection of its data.

METHOD

Subjects and Experimenters

The experiment employed 50 female subjects, all of whom were previously independently rated on hypnotic susceptibility on the Stanford Hypnotic Susceptibility Scale, Form A (SHSS; Weitzenhoffer & Hilgard, 1959). Only girls who had scored, respectively, in the upper and lower thirds of the Scale were used. The former were designated as high susceptibles (tranceable—T) and the latter as low susceptibles (untranceable—UT).

The study was conducted by three experimenters, one of whom selected subjects and randomly assigned them to the other experimenters and treatment conditions. For the latter, subjects were divided into five groups which differed only in the type of instructions they received in the second session of the experiment. The other two experimenters conducted the experiment proper, each testing half of the subjects, always without knowing the method of their selection or their susceptibility.

Instruments

Three performance tasks were employed in this experiment: tremor test, hand dynamometer, and weight endurance.

Tremor test. Stability of eye-hand coordination was measured by the subject's ability to hold a metal stylus in a hole without touching its edges. The stylus was attached electrically to a small box placed on a table at arm's length from the subject and at shoulder height. On the side of the box facing the subject was a brass plate with a hole $\frac{7}{32}$ inch in diameter in its center; whenever the stylus came into contact with the plate, an electrical circuit was completed so that an attached counter cumulatively registered one digit at each contact.

Hand dynamometer. A Stoelting Company pistol-grip dynamometer, calibrated in 1-kilogram units, was employed to measure grip strength.

Weight endurance. A $2\frac{1}{2}$ kilograms flat circular weight of the kind used in bar bells was taped and equipped with a strap which enabled it to remain stationary on the back of the subject's hand while her arm was outstretched. When held straight out from the shoulder, parallel to the floor, the weight could be brought in contact with a board which was connected to a microswitch so that very gentle pressure on the board closed an electrical circuit activating a Standard Electric Timer Company chronometer.

Instructions

The *base rate* and *exhortative* instructions of London and Fuhrer (1961) served as a model for the instructions used here. New instructions for this experiment were designed as variants of those. In all, five kinds of instructions were used:

Base rate instructions. These are minimal instructions in which the subject is simply told to do the task as well as possible; no special explanation or emphasis on good performance was included in this set of instructions.

Analgesia instructions. These instructions differ from base rate instructions in that subjects are additionally told to be insensitive to pain during their performance. The instructions read to the subjects are as follows:

> This entire project is primarily really an experiment on certain kinds of human potentialities. Psychologists have learned that most people tend to over-estimate their sensitivity to pain. This is especially true of women. Today, I'm going to ask you to do some simple physical tasks, just as you did the last time you were here. But this time, before we start, I'm going to tell you some things that will help you to minimize painful sensations and perform more efficiently even than you did previously.
>
> You are probably aware that most people use pain and fatigue as signs that they should stop what they are doing during all kinds of performances. But I wonder if you realize that, in so doing, they fail to realize that these things really function as nothing other than cues that they are approaching their maximum performances and that they can keep on going beyond that point. Perhaps more important, if they are aware of these facts and use them properly as cues, then intensifying their performance can make any pain seem less noticeable.
>
> Of course, in your experience you probably have not had much occasion to think about these facts, but now that you are aware of them in advance, and understand how crucial this information is, and how it applies to you, you will have a tremendous advantage in performing each task. You will feel neither pain nor discomfort to as great a degree, and as a result, will give a better performance than you ever have before.

Exhortative instructions. These instructions are almost identical with those employed by London and Fuhrer (1961), and consequently, only illustrative sections are included here.

> . . . I'm going to tell you something that will help you to discover the real extent of your abilities. It has been shown that thinking of a movement and making a movement are closely related. Similarly, really desiring to do something helps one to overcome any difficulties there may be in doing it . . . you will find that the more you desire to do it, the more you will think about succeeding . . . you will actually begin to feel stronger, steadier, and more capable, enabling you to perform at your best.

Combined analgesia and exhortative instructions. These instructions combine the content of analgesia and exhortative instructions, without noticeably increasing their length.

Hypnotic induction. The hypnotic induction procedure was identical with that used by London and Fuhrer (1961) and by Rosenhan and London (1963a, 1963b). In essence, it consists of the induction procedure of the Stanford Hypnotic Susceptibility Scale, Form A (SHSS; Weitzenhoffer & Hilgard, 1959), modified to suit the purposes of these experiments, which required that very susceptible and relatively unsusceptible subjects both be subjected to hypnosis. The modified portions included statements like:

> Just relax in the chair and make yourself very comfortable. That's fine. Now, just let your eyes close and let yourself rest in the chair—that's it, just leave your eyes gently closed while I talk. . . . The important thing is simply for you to relax in the chair and enjoy this experience while I say a few things that may be of interest to you. . . .
>
> Most people, when they undergo the kinds of hypnotic experiences which you have had here previously, wonder just how well they have been responding to the suggestions that are given. They are concerned about whether or not they have been performing as they were expected to and whether or not they are really good hypnotic subjects. Perhaps you have had these same questions in connection with your earlier experiences here. The fact is that your performance earlier indicated clearly, in some of its significant details, that you would be an excellent subject for the purposes of the experiment we are doing now. Because you *are* a good subject, we want you to participate in the hypnotic phase of this experiment, and I am going to prepare you, as you remain relaxed in your seat, for that phase. . . .
>
> As you relax there, very deep and very drowsy, you find it easy just to let things ride by your mind without attending very much to them. You can hear my voice clearly enough, and you can hear, in actuality, any other sounds that are around this room—but it is too much effort for you to pay attention to those outside sounds, so nothing makes much impression on your mind except the things that I am saying to you.
>
> In fact, your attention becomes so fixed on what I am saying that you can completely forget about anything else—you stop thinking about anything except what I am about to tell you—you listen to me, very carefully, without any real effort on your part, and you disregard all other thoughts and memories. You become conscious of my voice only, it fills your mind and thoughts and leaves room for you to attend only to what I say.
>
> In a few seconds, we shall begin working some tasks. You are already quite familiar with them, you have already performed all of them. But this time, throughout everything you do, you will not come out of this state until I tell you to. In just a second, I shall count slowly to three. When I do, your eyes will open and look straight ahead, but you will remain as deeply hypnotized as you are right now. One—two—three. Open your eyes. That's fine.

Procedure

The experiment was divided into two sessions held a week apart. In each session, the three performance tasks were always administered in the order: tremor test, hand dynamometer, and weight endurance. The subjects always performed the tasks first in the waking state and then in the hypnotic state, except for a single control group of 10 subjects (called hypnosis control) who repeatedly performed all the tasks in the waking state only.

Session I. After a few minutes for establishing rapport and familiarizing the subject with the situation, the three motor tasks were presented consecutively with 1 minute of rest between each task. First the tremor task was explained to the subject, and the stylus was placed in her hand. After leveling the position of the box with the subject's outstretched arm, the experimenter signaled her to begin. The task consisted of three trials, each 30 seconds long, with a 30-second rest period between trials. For the dynamometer, the subject was told to test her grip strength by squeezing the device as hard as possible when instructed, and to continue squeezing until told to stop. Five seconds were allowed for this task, then a 30-second rest period, after which the test was repeated. Throughout the experiment, the dominant hand was used for this test. For the weight endurance test, the subject was told to hold the weight up against a board for as long as she could, and it was explained to her that the timing device operated only when the weight actually pushed the board up. The test was discontinued as soon as the subject dropped her hand, bent her elbow, or otherwise gained support. The subject's nondominant hand was always used for this task. She was then given a 5-minute rest after which hypnosis was induced by verbatim instructions; the subject then repeated the tasks in the same order. The 10 subjects who served as the hypnosis control group, however, performed the tasks in the waking state both times.

Session II. The instructions for hypnosis control and for instruction control groups were the same as in Session I. Experimental groups were given special sets of instructions depending upon the experimental conditions to which they were assigned. The session began for the exhortation group, for example, with a strong motivating statement including a plea that their performance should surpass that of Session I because of their highly motivated condition. A further short exhortation was read prior to each task. The same procedure with appropriately different instructions was applied to the analgesia and combination groups.

After the final task, while the subject was still hypnotized, the hand analgesia item (Number 1, Form I) of the Stanford Profile Scales of Hypnotic Susceptibility (Hilgard, Lauer, & Morgan, 1963) was administered.[2] The subject was then awakened and asked to tell what she thought the design of the

---

[2] A flaw in the administration of this item invalidated its results, which are therefore not reported here.

TABLE 1

MEAN TREMOR SCORES FOR ALL GROUPS AND CONDITIONS

Condition	Control group		Experimental group		
	Hypnosis	Instruction	Analgesia	Exhortation	Combined analgesia & exhortation
Session I: control					
T					
Awake	49.4	68.8	78.2	38.4	62.0
Hypnosis	37.6	65.6	48.0	51.8	94.0
UT					
Awake	48.0	47.4	43.4	40.4	45.6
Hypnosis	38.2	57.2	37.6	29.2	48.4
Session II: experimental					
T					
Awake	40.0	68.4	59.4	24.4	70.4
Hypnosis	36.2	41.2	38.2	21.8	72.4
UT					
Awake	26.8	58.4	43.6	42.2	42.8
Hypnosis	26.2	52.2	28.5	28.6	37.8

Note.—$N = 50$.

experiment was and what her own reactions were to its various phases.

RESULTS

Scores were computed for each of the tasks as follows: for tremor, scores consisted of the sum of errors over three trials; for dynamometer, the subject was assigned a score equal to the total kilograms of pull exerted in both trials during each condition; the weight endurance scores consisted of the total time in seconds that the subject held the weight aloft without supporting or bending her arm. Mean scores for all groups and treatments are presented in Tables 1, 2, and 3.

The design of this experiment permitted several variables to be examined separately. These include the differences between susceptibility groups (T and UT), the effects of practice, and the influence of special instructions on performance in both the waking and hypnotic states.

TABLE 2

MEAN DYNAMOMETER SCORES FOR ALL GROUPS AND CONDITIONS

Condition	Control group		Experimental group		
	Hypnosis	Instruction	Analgesia	Exhortation	Combined analgesia & exhortation
Session I: control					
T					
Awake	51.8	50.4	54.8	51.2	54.2
Hypnosis	52.4	52.6	53.4	51.4	55.8
UT					
Awake	52.6	59.2	54.8	56.6	56.4
Hypnosis	53.2	57.6	56.0	52.2	52.6
Session II: experimental					
T					
Awake	53.4	56.4	58.8	55.2	58.4
Hypnosis	54.4	53.2	59.0	58.4	64.0
UT					
Awake	54.8	59.4	57.2	61.6	58.4
Hypnosis	58.2	54.4	57.0	61.6	57.2

Note.—$N = 50$.

TABLE 3

MEAN WEIGHT ENDURANCE SCORES FOR ALL GROUPS AND CONDITIONS

Condition	Control group		Experimental group		
	Hypnosis	Instruction	Analgesia	Exhortation	Combined analgesia & exhortation
Session I: control					
T					
Awake	101.8	77.8	83.2	104.6	67.2
Hypnosis	85.0	80.2	68.0	80.4	65.0
UT					
Awake	72.0	118.8	111.6	88.8	85.6
Hypnosis	58.4	78.4	94.7	66.2	58.5
Session II: experimental					
T					
Awake	99.2	71.4	78.8	102.8	81.4
Hypnosis	90.8	54.2	72.0	92.8	85.4
UT					
Awake	73.2	109.6	102.4	104.0	97.0
Hypnosis	63.0	69.2	76.6	101.8	95.6

Note.—$N = 50$.

Susceptibility Base Rates

In Session I, where no special instructions were given, 20 T and 20 UT subjects performed all three tasks in the waking state first, then in the hypnotic state. Differences in these base rate performances across groups were examined by means of t tests between the absolute scores of T and UT subjects on each task. While most differences between the two groups were insignificant, there was a trend for the UT group to do better than the T group on the tremor task in both waking and hypnotic states, and on the dynamometer and weight endurance tasks in the waking state. On the weight endurance task, moreover, the hypnotic score drops considerably more for the UT group than for the T group, relative to their respective waking scores ($t = 1.93$, $df = 38$, $p < .06$).

Practice Effects

Since hypnosis was invariably administered subsequent to waking performance in each session, it is possible for differences between hypnotic and waking performances within a session to be the result of practice (or fatigue) rather than of the different states which are assumed to be involved. This question was examined by comparing the performances of the 10 subjects who were tested only in the waking state with those of the 40 subjects who received both waking and hypnotic tests. The comparison was made for Session I only, since the insertion of special instructions in Session II made it irrelevant there. For this purpose, each subject received a score representing the difference between his hypnotic and waking performance (H-A) or first from second waking performance (A_2-A_1) in the first session. These difference scores were then compared by means of t tests between groups. For all tasks, there were no significant differences between the difference scores, i.e., hypnotic scores differed no more from waking ones than did the two waking scores from each other, though all scores tended to decline from one performance to another. Differences between the states may not be explained, therefore, as the result solely of practice.

Effect of Special Instructions

Thus far we have compared susceptibility base rate performances and have tested for practice effects. It remains to be asked whether performance is affected by special instructions. To answer this question, four two-way analyses of variance were performed (McNemar, 1962). Each subject's waking performance score in the first session was subtracted from her waking score in the special instructions session. The same procedure was applied to the score in the hypnotic conditions

TABLE 4

ANALYSIS OF VARIANCE ON TREMOR DIFFERENCE SCORES

Source	*df*	*MS*	*F*
Susceptibility (A)	1	2,387	4.165*
Instructions (B)	3	102	—
A × B	3	273	—
Within treatments	32	573	
Total			

Note.—Session II: Hypnosis − Session I: Hypnosis.
* $p < .05$.

of the base rate session and special instructions session. These difference scores were then grouped according to the subject's susceptibility rating (T or UT) and according to the instruction group (instruction control, analgesia, exhortation, combined analgesia plus exhortation) to which they were assigned. The analyses thus compared *difference* scores of T and UT groups for the various special instruction conditions, comparing waking and hypnotic performances separately.

The use of special instructions to facilitate performance is significantly effective when applied to performances obtained in the hypnotic state on dynamometer and weight endurance tasks (see Tables 5 and 6), but it is not effective in facilitating performance in the waking state on any of our three tasks.

As the main effects of the analyses of variance for the hypnotic performance scores were significant, t tests between the special instruction groups were then performed. On the dynamometer task, analgesic instructions tend to be slightly more effective than base rate instructions (instruction control group) ($t = 1.91$, $df = 38$, $p < .06$), but exhortation instructions are significantly more powerful than both base rate ($t = 3.5$, $df = 38$, $p < .01$) and analgesic instructions ($t = 2.04$, $p < .05$). The *combined analgesic and exhortation* instructions also differ significantly from base rate instructions ($t = 2.05$, $df = 38$, $p < .05$), but by less than exhortation, which is by far the most effective instruction. The combination of exhortation and analgesia, if anything, may reduce the effectiveness of exhortation alone, though this is uncertain.

TABLE 5

ANALYSIS OF VARIANCE OF DYNAMOMETER DIFFERENCE SCORES

Source	*df*	*MS*	*F*
Susceptibility (A)	1	75.6	3.34
Instructions (B)	3	181.0	8.00***
A × B	3	30.0	—
Within treatments	32	22.6	
Total	39		

Note.—Session II: Hypnosis − Session I: Hypnosis.
*** $p < .001$.

On the weight endurance task, simple effects tests reveal essentially the same results as on the dynamometer. Analgesic instructions do not differ significantly from base rate instructions, but exhortation instructions do ($t = 2.5$, $df = 38$, $p < .05$). On this task also, exhortation is significantly more effective than analgesia ($t = 3.0$, $df = 38$, $p < .01$). The combined instructions, however, appear here to be more effective than exhortation alone, but not significantly so.

T-UT Dimension in Session II

There are no significant differences between T and UT subjects in the second experimental session for either waking or hypnotic performances on the weight endurance and dynamometer tasks, though T subjects show a stronger trend than UT subjects towards performance increment on the dynamometer task. There is no significant interaction between susceptibility and experimental conditions.

On the tremor task, however, though there are no differences between the various instruction conditions, there is a consistent difference between the susceptibility groups, such that in the hypnotic state, T subjects improve significantly more than UT subjects from the first to the second experimental session regardless of the instructions delivered in the latter ($p < .05$, see Table 4).

TABLE 6

ANALYSIS OF VARIANCE ON WEIGHT ENDURANCE DIFFERENCE SCORES

Source	*df*	*MS*	*F*
Susceptibility (A)	1	.04	—
Instructions (B)	3	3,443.33	6.64**
A × B	3	661.33	—
Within treatments	32	518.60	
Total	39		

Note.—Session II: Hypnosis − Session I: Hypnosis.
** $p < .01$.

DISCUSSION

The most striking result of this experiment is the finding that positively motivating instructions facilitate task performance more under hypnosis than under comparable waking conditions. The same effect is seen on two of three experimental tasks, and it is of interest that both of the tasks involved are ones which require *maximum exertion* rather than refined or skillful activity for their successful completion. It is apparent, moreover, that the effectiveness of the instructions is related to their explicitly motivating character; straightforward exhortations were significantly more effective on both dynamometer and weight endurance tasks than were the more oblique statements that the subjects would feel no pain.

The relatively greater effectiveness of special instructions in the hypnotic than in the waking state is not explained by differences in the base rate levels of performance for waking and hypnotic conditions. On the weight endurance task, performances in these states do not differ significantly and on the dynamometer task, they are virtually identical. Inferior hypnotic performance in the base rate session would, of course, permit greater hypnotic performance increments in the experimental session. While this does not occur among all subjects combined, it does tend to be reflected within the low susceptibility group.

In comparing the results of the present experiment with previous research in this area, it must be remembered that the analyses performed here generally use relative improvement rates within conditions as the basis for comparisons of motivational sets and susceptibility groups, whereas other studies generally compare absolute performance across subject groups and psychological states. On both weight endurance and dynamometer tasks, however, in this experiment, as in previous ones (London & Fuhrer, 1961; Rosenhan & London, 1963a), UTs do absolutely better than do T subjects. Performance of these groups on the tremor test is more ambiguous, in comparison with previous experiments, but tends to be consistent with the group differences on the other tasks in this experiment—UTs tend to perform better than Ts in the base rate conditions, and T subjects improve significantly more as a consequence of hypnosis than do UT subjects. The main difference between tremor and the other tasks in this experiment is found in the lack of differential effects of special instructions on tremor performance.

Aside from the comparison of different instructional sets, the design of the present experiment involved three major changes from that of the previous experiments which used the same subject selection criteria, performance tasks, and general experimental procedures: (*a*) the sample was obtained from a different population of college students (Stanford instead of Illinois or Pennsylvania), (*b*) selection for hypnotic susceptibility status had been made independently of the experiment, sometimes as much as 2 years prior to it, and (*c*) experimenters were blind to the susceptibility status of their subjects. Under the circumstances, the consistency of results among all three experiments, considering that the first two were themselves obtained on different sex groups in different places by different experimenters, suggests that the results obtained have considerable generality. The present experiment thus reinforces two of the principal conclusions of the previous ones: (*a*) UT subjects manifest greater initial performance ability than T subjects on simple maximum performance strength and endurance tasks, regardless of sex. (*b*) The effects of hypnosis as such tend to be minimal on the performances of either group on these tasks, but seem to have slight differential effects which interact with susceptibility, such that T subjects are likely to show very slight improvements, and UT subjects very slight decrements, relative to their base rate waking performances, when they are required to perform the same tasks under hypnosis.

The special significance of the present experiment, however, comes from two features of its results which are not implied in previous work: (*a*) it demonstrates, in the differences between performances on tremor and the other tasks, the task specific effects of hypnosis more clearly than previous experiments have done, and (*b*) it demonstrates the interaction of hypnosis and instructional sets in a way which sheds new light on the motivational properties of hypnosis.

Hypnosis and Task Specificity

Rosenhan and London (1963a) corrected a methodological error in London and Fuhrer's use of the tremor test (1961) and found that, using male subjects, improvement in performance is significant only among T subjects and only in the hypnotic condition. The present experiment, using the same methodology but female subjects, obtained virtually the same result. Tremor not only differentiated susceptibility groups in this experiment, however, but unlike the other tasks, it was unaffected by any of the special instruction sets designed to facilitate performance. This result strongly supports the argument of Rosenhan and London (1963b) that "hypnosis is a situation with potentially variable effects which interact per susceptibility status *per measuring instrument*," and that it is consequently "improper . . . to generalize about its effects without consideration of the susceptibility of the *Ss* and the limited scope of the experimental tasks." Were weight endurance, for example, to have been the only instrument used in the experiment, it would have seemed reasonable to conclude that instructions were critical determinants of task improvement in hypnosis, whereas susceptibility was not. Were tremor to have been the only instrument, the opposite conclusion would have seemed in order. As it is, neither generalization is sufficient.

Hypnosis and instructions

It is impossible to tell, from our results, the extent to which the irrelevance of special instructions to performance on the tremor test applies to other tests of psychomotor coordination. But it is very clear that instructions are critical, in combination with hypnosis, for performance on at least two very fatiguing strength and endurance measures. The insertion of these experimental conditions, moreover, seems to be more important than the susceptibility of the subjects. What is puzzling about these results, nevertheless, is the fact that neither hypnosis nor exhortative instructions, taken alone, seem to be significantly effective in improving performance on these tasks. The minimal effect of hypnosis is quite consistent with the results obtained by London and Fuhrer (1961) on the same tasks, but they found exhortation to be quite effective without the addition of hypnosis, whereas we found its effects insignificant under those conditions.

At all events, the present results seem to flatly contradict the assertion of Barber and Calverley (1963) "It may be that hypnotic induction procedures and task motivating instructions both produce heightened suggestibility . . . effective in motivating subjects . . . to perform to the best of their ability on experimental tasks." If the experimental tasks are maximum performance measures rather than tests of hypnotic susceptibility, coordination, or learning, then there is now a considerable body of results which says that hypnotic induction has no effect by itself, and in our case, neither do task motivating instructions!

If anything, these results would suggest that the hypnotic condition somehow functions to make a subject more receptive than he has previously been to *certain kinds* of maximum performance instructions. The brief instruction to "do as well as you can" produces no great effect, and even the more complicated statement that the subject will experience no pain does not yield significant improvements. It is only the insertion of an exhortative appeal, with or without assurances of analgesia, that somehow parleys the effect of the hypnotic state so that performance is much improved. While we do not pretend to have an explanation at hand for this result, it does seem clear that hypnosis cannot be explained in any simple way as an independently motivating state.

The failure of the analgesia instruction to significantly affect performance is not necessarily a simple reflection on its content; our motive for using this content in the first place was that Roush (1951) found that analgesia instructions were significantly effective in her quite adequately designed experiment. In the present study, moreover, there was a trend in the direction of improved performance under analgesic instructions which was sometimes significant. It seems likely that the lack of effect may have resulted from some insufficiency in the length, intensity, or some other aspect of this instructional set. Differentiation of the content of instruction sets from other characteristics of them is an important subject for further research, which we are now conducting; pilot results suggest that analgesia is a very effective instruction under some circumstances.

REFERENCES

BARBER, T. X., & CALVERLY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of task motivating instructions and attitudes toward hypnosis. *Journal of Abnormal and Social Psychology*, 1963, **67**, 557–565.

EYSENCK, H. J. An experimental study of the improvement of mental and physical functions in the hypnotic state. *British Journal of Medical Psychology*, 1941, **18**, 304–316.

HADFIELD, J. A. *The psychology of power*. New York: Macmillan, 1924.

HILGARD, E. R., LAUER, L. W., & MORGAN, A. H. *Manual for Stanford Profile Scales of hypnotic susceptibility, Forms 1 & 2*. Palo Alto: Consulting Psychologists Press, 1963.

HULL, C. L. *Hypnosis and suggestibility: An experimental approach*. New York: Appleton, 1933.

LONDON, P., & FUHRER, M. Hypnosis, motivation and performance. *Journal of Personality*, 1961, **29**, 321–333.

MCNEMAR, Q. *Psychological statistics*. (3rd ed.) New York: Wiley, 1962.

NICHOLSON, N. C. Notes on muscular work during hypnosis. *John Hopkins Hospital Bulletin*, 1920, **31**, 89.

REIGER, C. *Der hypnotismus*. Stuttgart: Gustav Fischer, 1884.

ROSENHAN, D., & LONDON, P. Hypnosis: Expectation, susceptibility, and performance. *Journal of Abnormal and Social Psychology*, 1963, **66**, 77–80. (a)

ROSENHAN, D., & LONDON, P. Hypnosis in the unhypnotizable: A study in rote learning. *Journal of Experimental Psychology*, 1963, **65**, 30–34. (b)

ROUSH, E. S. Strength and endurance in the waking and hypnotic state. *Journal of Applied Physiology*, 1951, **3**, 404–410.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Forms A & B*. Palo Alto: Consulting Psychologists Press, 1959.

WILLIAMS, G. W. A comparative study of voluntary and hypnotic catalepsy. *American Journal of Psychology*, 1930, **42**, 83–95.

YOUNG, P. C. An experimental study of mental and physical functions in the normal and hypnotic states. *American Journal of Psychology*, 1925, **36**, 214–232.

YOUNG, P. C. An experimental study of mental and physical functions in the normal and hypnotic states: Additional results. *American Journal of Psychology*, 1926, **37**, 345–356.

(Received February 17, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 47–51

THE INFLUENCE OF PERCEIVED MENTAL ILLNESS ON INTERPERSONAL RELATIONS[1]

AMERIGO FARINA AND KENNETH RING
University of Connecticut

This study was carried out to measure the influence of general attitudes toward mental illness on 2 aspects of interpersonal interaction: an S's performance on a motor task requiring cooperative effort and the perception of another after having interacted with him on the task. Ss were undergraduates who were induced to believe their co-worker was either "normal" or had been mentally ill. The results indicate that perceiving the co-worker as mentally ill is associated with better task performance, possibly because of the greater threat posed by a successful peer in comparison to that posed by a person believed to be maladjusted and inadequate. It was also found that when a co-worker is viewed as mentally ill, Ss prefer to work alone rather than with him and blame him for inadequacies in the joint performance even though objective measures do not justify these responses.

It seems clear that, in America at least, there is a generalized highly unfavorable attitude toward the mentally ill (Nunnally, 1961; Nunnally & Kittross, 1958). Evidence suggests that such a pandemic attitude can color the perception of a person believed to be mentally ill even when his behavior is, by all objective standards, "normal." For example, it has been found (Jones, Hester, Farina, & Davis, 1959) that subjects listening to standardized taped "interviews," dislike the speaker more when they are told (by the experimenter) that she is maladjusted than when they believe she is well adjusted.

Interesting as such studies are, in general they provide no more than a glimpse into prevalent American attitudes and stereotypes about the mentally ill. Conceivably, such attitudes are maintained through what Newcomb (1947) has called "autistic hostility." That is, these attitudes resist change because an individual holding them ceases to communicate with the class of persons unfavorably evaluated. Certainly, most Americans have very limited face-to-face contact with the mentally ill. The situation is reminiscent of the early studies on stereotypes (e.g., Katz & Braly, 1933) where it was shown that, in the main, Turks were strongly disliked in spite of the fact that few of the respondents had ever known one.

In the present study we are concerned with the modifiability of such stereotypes as a function of face-to-face interaction with a "mentally ill" person whose behavior in no way supports the common attitude toward him. We are asking, in effect, if the mentally ill person is a prisoner of his own reputation. In this, the present research is relevant to certain hypotheses recently suggested by Goffman (1963) regarding the stigma attached to the exmental patient.

In addition to studying the effect of direct interaction (something that earlier studies lacked) on the perception of a person believed to be mentally ill, the effects of such a belief on the performance of a dyadic cooperative task is also of interest. If, as the previously cited research would suggest, unfavorable attitudes toward the mentally ill are highly generalized, one would expect them to influence behaviors over a wide variety of interpersonal contexts. For this reason, in the interaction studied, a motor task which seems completely unrelated to mental illness was deliberately chosen.

The question now arises: what is likely to be the effect of believing a co-worker to be mentally ill on the performance of a cooperative task? There seem to be several possibilities. First of all, on the basis of an anxiety-performance decrement hypothesis, it might be expected that perceiving the co-worker as emotionally disturbed would lead to decreased adequacy of performance. This would occur

---

[1] This study was, in part, financed by a research grant (M-6167) from the National Institute of Mental Health, United States Public Health Service.

if the perceiver's level of anxiety were raised by the belief that the co-worker was, because mentally ill, unpredictable, tense, and dangerous. This increased anxiety should interfere with the performance on the very difficult task used.

On the other hand, two other possibilities suggest that seeing the co-worker as mentally ill would increase performance adequacy. It might be *less* threatening to work with someone who is seriously maladjusted and consequently in a poor position to evaluate others than to work with a normal person who might be more justifiably critical and judgmental. In this case, anxiety should be higher and, therefore, the performance lower when the co-worker is perceived as normal than when he is perceived as mentally ill. The same relationship is suggested by the very different possibility that when the co-worker is perceived as mentally ill, he is thought to be less competent to perform the task. The subject may compensate for the imagined inadequacy of his co-worker by trying harder. Therefore, better performance would be associated with the perception of the partner as mentally ill. In view of the seeming reasonableness of these contradictory alternatives and the absence of any guiding data, no predictions were made. It was thought, however, that if differences in adequacy of performance in fact emerged, their meaning would be clear when examined in the light of measures of perception of the co-worker.[2]

METHOD

Subjects. The subjects of this study were 60 male undergraduate students enrolled in psychology classes at the University of Connecticut. They were all volunteers and were given no pay for participating in the experiment. The subjects participated in pairs composed in all cases of 2 students who were unacquainted with each other.

Procedure. After the subjects were introduced, they were asked to sit at two desks facing each other. However, a partition was placed between the desks in such a way that the students could not see each other but could at all times be seen by the experimenter. They were then told that this study sought to determine how much a person could tell about another from certain standard cues and were instructed to refrain from talking to each other. It was explained that the experimenter wanted them to know some specific things about each other and that they were to write that information on blank sheets of paper which would subsequently be exchanged. They were asked to list three kinds of information: (*a*) the sort of person they were, (*b*) anything unusual about themselves, and (*c*) something about their plans for the future. They were urged to be as candid as possible inasmuch as it would be very important to the study.

When a student indicated he was finished (generally 5 to 15 minutes) his background information sheet was picked up by the experimenter and the subject was given a brief one-page questionnaire about his school and family. The sole purpose of this questionnaire was to facilitate the substitution of faked background information sheets for the real ones. These fake sheets, which described the writer either as normal or maladjusted, were placed on the students' desks while they were engaged in completing the questionnaire. They were then told that, as soon as they had completed the questionnaire, they were to read their co-worker's information sheet very carefully.

The normal information sheet was as follows:

> I tend to think of myself as a relatively normal person; at least I don't have what you could call any "problems." I enjoy going to college, but like to have my fun, too. I think I'm pretty popular with my group, am engaged, and am doing pretty well (29 QPR)[3] in school.
>
> Frankly, I can't think of anything that's "unusual" about myself.
>
> As for my goals for the future, after graduating from college, I plan to get married and hope to go to graduate school.

The "sick" one, on the other hand, read:

> You asked us to be candid, so here goes: I have certain problems in adjustment which I first noticed in high school and which still bother me quite a bit. I guess I am somewhat different from most people. I tend to keep pretty much to myself and, frankly, I don't really have any close friends. At school I am doing pretty well (29 QPR).
>
> I suppose what's most unusual about me is that twice (once in my senior year in high school and once in college), I have been placed in a mental institution when I had a kind of nervous breakdown.
>
> As for my goals for the future, after graduating from college, I hope to go to graduate school.

When both subjects had finished, they were taken to another section of the room where a toy manufactured in Sweden and sold under the brand name *Labyrintspel* (Labyrinth Game) was on top of a table. Subjects were asked to sit on chairs in front of the *Labyrintspel,* their chairs were drawn up against each other, and they were asked to examine the toy. It resembles a box, measuring 13 × 11 inches, it is approximately 4¾ inches high, and there is a knob at each of two sides. Each knob controls an angle of the plane at the top of the box so that a small steel ball on it can be made to follow any path desired by appropriate and simultaneous manipulation of the two knobs. The movable plane is a maze formed by raised wooden partitions and a series of 40 holes. There is only one possible path from the starting point to the goal and that is indicated by a black line. There are 59 points at which the ball can fall into a hole and these are numbered from 1, which is the first pitfall, to 59, which is the last, before the goal.

---

[2] It is of course conceivable that two of these processes might counterbalance each other in which case no differences would emerge in spite of the presence of these processes.

[3] About a B average.

The subjects were each assigned a knob and they were told to make the ball go as far as possible on each trial and when the ball fell in a hole they were to pick it up as it emerged, place it in the starting box, and begin another trial.[4] They were given three practice trials and told that they were free to talk about the task at any time. The performance consisted of 50 consecutive trials with a 1-minute rest period after Trial 25. For each trial the hole in which the ball fell was noted and the time required was also recorded. The latter variable was not analyzed and served primarily to reduce the time the experiment required by causing the subjects to work more rapidly.

At the termination of the maze performance subjects returned to their desks and were given a final questionnaire to complete. They were told that the questionnaire was strictly confidential and would not be seen by their co-worker. It consisted of 14 statements which subjects could answer by placing a check mark at the appropriate place and a fifteenth statement which called for a free description of the co-worker. After the questionnaire, a discussion was held with the subjects in order to determine if any were suspicious and, subsequently, to reveal the true nature of the study to them. Two subjects guessed the real intent of the study and were discarded.

Experimental design. The subjects were randomly assigned to one of three groups, each group being composed of 10 pairs of subjects for a total of 30 pairs. The groups differed only with respect to the information sheet received. In the normal-normal (N-N) group, both subjects received a normal information sheet. For the normal-sick (N-S) group one student received the normal information sheet but the other received the sick one. The last group was the sick-sick (S-S) group and each student in the group received the sick background information sheet. For the analysis of the performance measures, these three groups are compared. However, for the measures of perception only two groups of 30 subjects each were compared. Those who received the sick background information sheet comprised one while those receiving the normal sheet formed the other group. The three groups were combined into two after it was determined that the perception of the co-worker was not influenced by the background information sheet the co-worker received.

RESULTS

An important fact to determine at the outset is how successful the manipulation was in inducing the subjects to perceive their co-worker in line with experimental intentions. On the final questionnaire there were four items which are relevant to this question. Those who received the sick sheet, in comparison to those who received the normal one, described their co-worker as: less able to get along with others, less able to understand others, less able to understand himself, and more unpredictable. For these four items the differences reached significance levels ranging from $p < .01$ to $p < .0001$. These results indicate that the manipulation was completely successful and, in addition, confirm Nunnally's (1961) findings that mentally ill persons are perceived as unpredictable.

The results of the performance on the *Labyrintspel* are presented in Table 1. The numbers in this table represent sums of scores obtained for the 50 trials. The N-N group was least adequate while the S-S group performed best, suggesting that perceiving the co-worker as mentally ill enhances performance. These results were analyzed by means of a trend analysis (Edwards, 1960) which showed that only the trials main effect reached a statistically significant level. Neither the groups main effect nor the Groups × Trials interaction was found to be statistically significant. Since it had been expected that the N-N and S-S group scores would be most disparate (whatever the effect of perceived mental illness), an additional analysis was carried out. Here the performance of the N-N and S-S groups was compared by means of a t test. The results indicate that the S-S group performance was reliably better than that of the N-N group ($t = 2.2$, $df = 18$, $p < .05$) and suggest that perceiving the co-worker as mentally ill facilitates performance.

Analysis of questionnaire items concerned with social perception (other than those mentioned at the beginning of this section) revealed that the background information sheet

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TOTAL SCORE OBTAINED FOR 50 TRIALS ON THE LABYRINTSPEL

Group	*M*	*SD*
N-N	252.3	71.8
N-S	275.5	101.8
S-S	334.8	94.6

---

[4] A small percentage of subjects (15% overall) had had some prior experience with the toy but the percentage did not differ across groups.

TABLE 2

NUMBERS OF SUBJECTS DESCRIBED AS HELPING OR HINDERING PERFORMANCE AS A FUNCTION OF BACKGROUND INFORMATION SHEET RECEIVED

		Co-worker described as: Helping	Co-worker described as: Hindering	χ^2	df	p
Background information sheet received	normal	27	3	5.7	1	.02
	sick	18	12			

had a significant influence on the perception of the co-worker in spite of the opportunity the students had of interacting with him. The results from one item which required the subjects to state whether they thought their co-worker helped or hindered in the performance of the task are presented in Table 2. Subjects who received the sick information sheet, more often than those receiving the normal sheet, reported their co-worker hindered the joint performance. This occurred *in spite of the fact* that the former group did objectively better on the task than the latter. Subjects were also asked to indicate whether they preferred working on the task by themselves (controlling both knobs) or with their partner. The relevant data are presented in Table 3. They clearly support the idea that believing an individual to be mentally ill reduces the perceiver's willingness to work with him. A third item of this kind was in the same direction as the preceding two in that the sick co-worker tended to be held more responsible for causing the ball to drop into a hole. However, the difference between the groups did not reach statistical significance.

Of the remaining items on the questionnaire, those concerned with estimates of the co-worker's intelligence and his ability to perform well on similar motor tasks showed no difference between groups. While sick co-workers were judged to be less liked by *others* than normals ($p < .01$), subjects themselves did not rate a sick co-worker any less likable than a normal one. Neither did subjects feel any less "comfortable" working with a sick co-worker compared to a normal one. These two findings seem to be at variance with the data already presented which would lead one plausibly, if not logically, to expect differences on these items. Also somewhat surprising is the absence of any difference between groups on questions dealing with how much attention the co-worker gave to the task and how willing the subject would be to work with his partner on a future, similar task.

DISCUSSION

The most important conclusion of this study is that believing an individual to be mentally ill strongly influences the perception of that individual; this is true in spite of the fact that his behavior in no way justifies these perceptions. When a co-worker is viewed as mentally ill, subjects prefer to work alone rather than with him on a task and also blame him for inadequacies in performance. Since objective measures of performance do not warrant such responses, these findings attest to the importance of believing another to be mentally ill as a factor in interpersonal relationships.

The significance of this variable is further emphasized by several considerations. The subjects of this study were young and well educated, characteristics shown to be associated with relatively favorable attitudes toward the mentally ill (Nunnally, 1961). Also, it seems very likely from the data referred to at the end of the Results section that the sick

TABLE 3

NUMBERS OF SUBJECTS CHOOSING TO WORK WITH PARTNER OR ALONE AS A FUNCTION OF BACKGROUND INFORMATION SHEET RECEIVED

		Choice to work with: Partner	Choice to work with: Alone	χ^2	df	p
Background information sheet received	normal	17	13	7.1	1	.01
	sick	6	24			

sheet either created a rather strong sympathy reaction in the subjects or made them less willing to say anything unfavorable about the co-worker. In spite of these counteracting factors it was still found that the generalized negative attitude had a strong effect on interpersonal relationships.

The finding of a difference in performance on the *Labyrintspel* suggests that perceiving someone as mentally ill affects a very broad spectrum of interpersonal behaviors. That is, while it might be expected that such a perception would influence a person's decision to hire someone as a baby sitter, the behavior measured in this study seems quite peripheral to mental illness. What factors may account for this finding?

It will be remembered that two possibilties were suggested for anticipating the finding obtained: (*a*) one was based on the hypothesis that a normal and successful peer would be threatening and (*b*) another assumed that compensation for perceived inadequacy would occur. If the first hypothesis were true, it would imply a decrement in performance in the N-N group while the second would suggest a facilitation of performance in the S-S group. Obviously these two hypotheses are not mutually exclusive.

It was decided to test the second of these hypotheses by running an additional group of 10 pairs of students. In this group, the clumsy-clumsy (Cl-Cl) group, each student received a background information sheet which was identical to the normal one except for the addition of one phrase: "About the only thing I can think of is that I'm 'all thumbs'; I don't learn to do things with my hands very well." The success of the manipulation was checked by comparing the N-N to the Cl-Cl group on the item which asked for an estimation of the co-worker's dexterity at motor tasks such as the one they had just completed. The z score yielded by the Mann-Whitney U test was 4.1 which is highly significant and in the expected direction. However, the task performance of the group was not different from that of the N-N group nor were the groups different with respect to their perception of the co-worker. These results are inconsistent with the compensation hypothesis, leaving the threat hypothesis as a possibility.

Of course, it could well be that the list of alternative hypotheses which could explain the original S-S—N-N difference has not been exhausted. For example, Ghiselli and Brown (1955, pp. 293–294) cite evidence indicating that performance on certain tasks is improved by distracting influences such as noise. Perceiving the co-worker as mentally ill may constitute a distracting influence and may have enhanced performance. Such hypotheses, clearly, may be subjected to direct test.

It has been demonstrated that, under certain circumstances, what a person reportedly says about himself significantly influences the interpretation of his behavior by another even though the behavior does not justify that interpretation. But, it might be argued, the type of encounter studied here was brief and superficial. If contact had been prolonged and intimate, isn't it reasonable to suppose that one's interpretation of another's behavior would be based less on stereotypes and more on the behavior itself? This is possible. Insofar as stereotypes about mental illness are concerned, however, this objection can be answered by submitting that it is precisely because of these stereotypes that distortions in perception are likely to occur in the *initial* phases of interaction and thus reduce the likelihood of further interaction of the kind necessary to eradicate such stereotypes.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.

GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1955.

GOFFMAN, E. *Stigma*. Englewood Cliffs, N. J.: Prentice-Hall, 1963.

JONES, E. E., HESTER, S. L., FARINA, A., & DAVIS, K. E. Reactions to unfavorable personal evaluations as a function of the evaluator's perceived adjustment. *Journal of Abnormal and Social Psychology*, 1959, **59**, 363–370.

KATZ, D., & BRALY, K. Racial stereotypes of 100 college students. *Journal of Abnormal and Social Psychology*, 1933, 28, 280–290.

NEWCOMB, T. M. Autistic hostility and social reality. *Human Relations*, 1947, **1**, 69–86.

NUNNALLY, J. C., JR. *Popular conceptions of mental health*. New York: Holt, Rinehart & Winston, 1961.

NUNNALLY, J. C., JR., & KITTROSS, J. M. Public attitudes toward mental health professions. *American Psychologist*, 1958, **13**, 589–594.

(Received February 28, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 1, 52–69

THE PROBLEM OF VALIDITY IN FIELD STUDIES OF PSYCHOLOGICAL DISORDER[1]

BRUCE P. DOHRENWEND[2] AND BARBARA SNELL DOHRENWEND

Columbia University *City College of the City University of New York*

Results of over 25 attempts to count untreated cases of psychological disorder in community populations are reviewed, and the problem of validity in the measures of disorder is analyzed. Evidence of validity is found to be scant. The position is taken that, with no generally accepted criteria available, and no universe of content agreed upon, construct validity takes on added importance. A lead to 1 possible nomological net for the construct of psychological disorder is the consistent finding that the lowest socioeconomic stratum has the highest rate of symptomatology. This lead is developed with reference to the transience of symptomatology found in extreme situations, in contrast to the persistence of symptomatology observed in patients and in studies of experimental neurosis.

For over 100 years, researchers have been reporting relations between sociocultural factors and psychological disorder. An early instance is Jarvis' finding, in 1856, that the "pauper class" in Massachusetts furnished proportionally 64 times as many cases of "insanity" as the "independent class (Sandifer, 1962)." Like Jarvis, most later investigators have tried to answer questions about psychological disorder in different groups by examining the characteristics of mental patients. A classic example is the research of Faris and Dunham (1939) on the distribution of admissions to mental hospitals from Chicago which, replicating Jarvis' results, showed a heavy concentration of cases from the central slum sections. The investigators' interpretation was that social disorganization might be causing psychological disorder.

Dunham (1961) and others (e.g., Felix & Bowers, 1948; Gruenberg, 1955) have, however pointed to the difficulty of interpreting the results of studies in which the fact of being in treatment constitutes the sole definition of disorder. Treatment rates vary, for example, with the availability of treatment facilities, and with public attitudes towards their use. Either could be responsible for spurious relations between sociocultural factors and rates of illness measured by number of cases in treatment.

FIELD STUDIES

Recognition of such problems has stimulated more than 25 attempts to count untreated as well as treated cases of psychological disorder. Excluded from this tally are studies which have reported scores measuring symptomatology, but have not indicated what is to be considered a case (e.g., Gurin, Veroff, & Feld, 1960; Langner, in press).

Typically, the whole population in a specified geographical area has been included. In the few investigations relying on sample estimates, the *n*'s have usually been large; for example, probability samples of 1,660 in the Midtown Study (Srole, Langner, Michael, Opler, & Rennie, 1962), and 1,010 in the Stirling County study (D. C. Leighton, Harding, Macklin, Macmillan, & Leighton, 1963).

In all but two studies, the rates we have extracted represent prevalence during a period of a few months to a few years. The two exceptions, a study by Fremming (1951) and a summary of 18 small studies by Strömgren (1950), counted as one study here, present lifetime prevalence rates.

The overall results are shown in Table 1. The first thing to note is the range in rates:

---

[1] This work has been supported in part by Grants MH 07327-01 and 07328-01 from the National Institute of Mental Health, United States Public Health Service. We are grateful to Jack Elinson, Lawrence C. Kolb, and Thomas S. Langner for their helpful criticism of this paper, and to Robert J. Smith for his help with Japanese census data.

[2] Community Psychiatry Research Unit, Columbia University, 128 Ft. Washington Avenue, New York 32, New York.

TABLE 1

PERCENTAGE OF PSYCHOLOGICAL DISORDER REPORTED ACCORDING TO GEOPOLITICAL AREA AND RURAL VERSUS URBAN STUDY SITE

Site	North America	Northern Europe	Asia	Africa
Rural	1.7 (Eaton & Weil, 1955) 1.9[a,b] (Rosanoff, 1917) 6.9 (Roth & Luton, 1943) 18.0 (Trussell et al., 1956) 50.0+[c] (D. Leighton et al.[d], 1963) 64.0 (Llewellyn-Thomas[d], 1960)	1.1[a] (Kaila, Study II) 1.3 (Brugger, 1931) 1.3 (Kaila, Study I) 3.5[a] (Brugger, 1937) 4.2 (Strömgren, 1950) 7.5 (Brugger, 1933) 9.0 (Mayer-Gross, 1948) 11.9 (Fremming, 1951) 13.2 (Primrose[d], 1962) 13.6[a] (Essen-Möller, 1956) 23.2 (Bremer[d], 1951)	0.8 (Uchimara et al., 1940) 1.0 (Lin, small town, 1953) 1.1 (Lin, village, 1953) 2.7[a] (Akimoto et al., 1942)	40.0 (A. Leighton et al.[d], 1963)
Urban	1.8[a] (Lemkau et al., 1942) 2.3 (Cohen et al., 1939) 3.4 (Manis et al.[d], 1964) 10.9 (Pasamanick et al., 1959) 23.4 (Srole et al.[d], 1962) 32.0[a] (Cole et al.[d], 1957)	1.0[a] (Kaila, Study I) 1.1[a] (Kaila, Study II) 15.6[a] (Gnat et al.[d], 1964) 33.0 (Taylor & Chave[d], 1964)	1.1 (Lin, city, 1953) 3.0[a] (Tsuwaga et al., 1942)	45.0 (A. Leighton et al.[d], 1963)

[a] Calculated by B. S. Dohrenwend.

[b] Includes some urban districts not reported separately.

[c] ". . . our conclusion from all the available information is that at least half of the adults in Stirling County are *currently* suffering from some psychiatric disorder defined in the American Psychiatric Association *Diagnostic and Statistical Manual* [p. 356]."

[d] Age range limited.

from less than 1% to over 60%. What factors could conceivably account for variation of such magnitude?

Table 1 groups the studies according to the geopolitical area in which they were conducted, and according to whether the study site was rural or urban. These contrasts in setting do not seem to account adequately for the variation in rates. In the relatively few studies done in Asia, the rates are very low, suggesting the possibility of a difference between Western and Asian populations. No such difference is indicated, however, by the two figures reported for Africa, which are near the top of the range found in North American and northern European studies. Moreover, within North America and northern Europe, there is no evidence that the study site is associated with the rate of disorder, since almost the full range is represented within each of four categories: North American rural, North American urban, northern European rural, and northern European urban.

Another possible explanation of differences in rates is the age range included in the study. As indicated in Table 1, 11 studies covered a limited age group. In all of these, the youngest ages, in which the minimum rate is usually found (see Table 3), are excluded. With one exception, the rates reported for these age-restricted studies range from moderate to high. The single exception, the study by Manis, Brawer, Hunt, and Kercher (1964), includes the same range as the Midtown Study by Srole et al. (1962), 20 through 59 years. The low rate reported by Manis et al. underlines the fact that elimination of the youngest age group is not a sufficient explanation of high reported rates.

A number of critics have argued that the factors primarily responsible for variation in rates are methodological (e.g., Blum, 1962; Pasamanick, 1962; Plunkett & Gordon, 1960). One such factor is the thoroughness of the data collection procedures. Investigators who collected data directly from the subjects, rather than relying entirely or partially on indirect sources of information, might be expected to detect more cases, as has happened with physical illness (Cartwright, 1957). On the other hand, still more thorough investigation through physical examinations might reveal organic bases of symptoms, thereby reducing rates below those reported when subjects' self-descriptions are the investigator's only source of information. Still another methodological factor which might affect rates is the conception of what constitutes a case. This conception may, as Szasz (1961) implied, have changed with time, expanding and becoming more inclusive over the years.

The effects of these methodological factors are tested in Table 2 by classifying the studies according to thoroughness of data-collection procedures and decade of publication. There is a general tendency for more direct contact with subjects to produce higher rates. The importance of direct contact is given further weight by the work of authors in italics in Table 2, i.e., Cole, Branch, and Orla, 1957; Eaton and Weil, 1955; Rosanoff, 1917; and Roth and Luton, 1943. In each of these studies, rates found in the same population are higher among subjects with whom more direct contact was made. There is, however, no

TABLE 2

PERCENTAGE OF PSYCHOLOGICAL DISORDER REPORTED ACCORDING TO THOROUGHNESS OF DATA COLLECTION PROCEDURES AND DECADE OF PUBLICATION OF STUDY

Decade of publication	Directness of contact with subjects			
	Indirect contact: records and/or informants	Partial direct contact: records and/or informants, and interviews with some subjects	Direct contact	
			interviews with all subjects	Interviews and physical examination of all or most subjects
1910–1919		*1.3 (Rosanoff, 1917)*	*3.6 (Rosanoff, 1917)*	
1920–1929				
1930–1939	2.3 (Cohen et al., 1939) 4.2 (Strömgren[b], 1950)	1.3 (Brugger, 1931)	3.5[a] (Brugger, 1937) 7.5 (Brugger, 1933)	
1940–1949	1.1[a] (Kaila, Study II) 1.8[a] (Lemkau et al., 1942) 9.0 (Mayer-Gross, 1948)	0.8[a] (Uchimara et al., 1940) 1.2[a] (Kaila, Study I) *6.4 (Roth & Luton, 1943)*	2.7[a] (Akimoto et al., 1942) 3.0[a] (Tsuwaga et al., 1942) *12.4 (Roth & Luton, 1943)*	
1950–1959	*1.2 (Eaton & Weil, 1955)*	*2.3 (Eaton & Weil, 1955)* 11.9 (Fremming, 1951) *28.0[c] (Cole et al., 1957)*	1.1 (Lin, 1953) *2.9 (Eaton & Weil, 1955)* 13.6[a] (Essen-Möller, 1956) *35.4[c] (Cole et al., 1957)*	10.9 (Pasamanick et al., 1959) 18.0 (Trussell et al., 1956) 23.2 (Bremer, 1951)
1960 and after			3.4 (Manis et al., 1964) 15.6[a] (Gnat et al., 1964) 23.4 (Srole et al., 1962) 33.0 (Taylor & Chave, 1964) 41.0 (A. Leighton et al., 1963) 50.0+[d] (D. Leighton et al., 1963)	13.2 (Primrose, 1962) 64.0 (Llewellyn-Thomas, 1960)

[a] Calculated by B. S. Dohrenwend.

[b] Eleven of the small studies summarized in Strömgren's rate were published in the 1930s, seven in the 1920s.

[c] Rates for wives are calculated for direct contact and other family members for partial contact on the basis of the statement that wives were usually interviewed; the base for each rate is the number of families interviewed.

[d] ". . . our conclusion from all the available information is that at least half of the adults in Stirling County are *currently* suffering from some psychiatric disorder defined in the American Psychiatric Association *Diagnostic and Statistical Manual* [p. 356]."

evidence that physical examination lowered the reported rates of psychological disorder. The rates yielded by the five studies which included physical examinations cover most of the range.

There is a clear tendency for rates to be higher in more recent studies. Especially notable is the increase in the 1950s and 1960s. It could be argued, of course, that rather than observing the effects of changes in the researchers' concepts, we are observing a consequence of the times we live in. However, in view of statements suggesting a shift from exclusive concern with avoiding overestimation (e.g., Cohen, Fairbank, & Greene, 1939, p. 113; Lemkau, Tietze, & Cooper, 1941, p. 635; Rosanoff, 1917, p. 137) to concern with complete enumeration (e.g., Bremer, 1951, p. 12; D. C. Leighton, et al., 1963, p. 195), it seems premature to infer a true change in prevalence between the 1940s and 1950s. Moreover, there is ample evidence from studies in which both inclusive and relatively exclusive standards were applied to the same populations that rates can be markedly affected by these standards. Essen-Möller (1956, p. 95), for example, reported a rate of 13.6% for "diagnoses constituting the main subject of most psychiatric population studies," but an average of 54.7% (calculation by B. S. Dohrenwend) for whom pathology was not definitely absent. A similar contrast is offered by the figures 23.4% in the impaired group, and 81.5% judged less than "well" in the Midtown Study (Srole et al., 1962, p. 138). Another comparison with the Midtown Study is provided by a recent investigation (Manis et al., 1964) which used 22 symptom items taken from the Midtown interview schedule. While obtaining a distribution of responses similar to that reported by the Midtown researchers, Manis and his colleagues found that their decision to include only severe psychological disorder yielded a rate of 3.4%.

The continuing importance of methodological problems is emphasized by the difficulty of comparing results even from studies carried out by the same investigators. For example, the Leightons and their colleagues (A. H. Leighton, Lambo, Hughes, Leighton, Murphy, & Macklin, 1963) concluded, concerning their own studies: "the differences and similarities between the Yoruba and Stirling figures are to an unknown degree under the influence of differences in the procedures employed in the two studies, [p. 124]." To complicate the problem of comparison further, various procedures used within one of these, the Stirling County study, yielded different rates of psychopathology (D. C. Leighton et al., 1963, pp. 123, 127).

PROBLEM OF VALIDITY

Given the variability of both procedures and results in attempts to assess "true" prevalence, the salient question is which, if any, among these field studies has produced valid measures of psychological disorder. Despite their deficiencies, studies restricted to treated disorder had a clear advantage. The fact that a person is in treatment usually indicates that he cannot function unaided in his customary social environments. The clinician diagnosing a patient has a "presenting problem" with which to start, so that the question he must answer is not *whether* something is wrong, but rather *what* is wrong. The diagnostic result of this analysis, moreover, can be changed on the basis of repeated observations and interviews over a course of treatment.

The investigator of untreated disorder must work without the aids to diagnosis inherent in the clinical setting. Evidence of the difficulties he faces are found in the results of the psychiatric screening attempts associated with Selective Service in the United States during World War II. The psychiatric judgments were extremely unreliable, rejection rates within the same region varying in a number of areas by a factor of three to one (Star, 1950b, pp. 552, 554). Moreover, there is no evidence that among the unreliable judgments one or another was more effective, since strictness of screening procedures bore little relation to subsequent rates of separation on psychiatric grounds (Ginzberg, Anderson, Ginsburg, & Herma, 1959b, Ch. 11).

Nevertheless, clinical judgment is the tool relied upon for case identification in almost all studies which include untreated as well as treated psychological disorder. In most, psychiatric diagnoses are made, and findings are presented in terms of categories such as those described in the *Diagnostic and Statistical Manual* of the American Psychiatric Association (1952). Neither the information available to the judge, nor the criteria on which the diagnoses were based are usually reported in detail in these investigations. The validity of the results is assumed to be implicit in the diagnostic process, a shaky assumption in light of the World War II experience with psychiatric screening.

A few investigators, recognizing the difficulty of placing untreated cases in diagnostic categories, have also made judgments in more general terms, such as probability of pathology (Essen-Möller, 1956; Rosanoff, 1917), likelihood of being psychiatric cases (A. H. Leighton et al., 1963; D. C. Leighton et al., 1963), or degree of severity ranging from "well" to "incapacitated" (Langner & Michael, 1963; Srole et al., 1962). Among these relatively sophisticated investigations, the Leightons' Stirling County and Nigeria studies, and the Midtown Study by Srole and his colleagues stand out. Although they differ in several aspects of their assessment procedures, they share an important innovation. Both used structured questionnaires, thereby providing a standard, explicit set of data for psychiatric assessment. In this respect, these studies represent the methodologically most advanced epidemiological investigations of untreated and treated psychological disorder.[3] The question is whether procedures of the Midtown, Stirling County, and Nigeria studies have dealt adequately with the central methodological problem of validity.

Content Validity

Content validity involves a demonstration that the items used are a representative sample from a universe generally accepted as defining the variable to be measured (Cronbach & Meehl, 1955). In the Midtown Study, the universe from which items were drawn was defined by Srole et al. (1962) as "the most salient and generalized indicators of mental pathology [p. 60]." Behavioral scientists selected a group of items from the Army Neuropsychiatric Screening Adjunct (Star, 1950a) and the MMPI (Dahlstrom & Welsh, 1960) "consisting principally of the psychophysiologic manifestations and those tapping the anxiety, depression, and inadequacy dimensions [Srole et al., 1962, p. 42]." In addition, the psychiatrists independently contributed 40 items "bearing particularly on psychosomatic symptoms, phobic reactions, and mood [p. 60]." The final decision determining the 120 items actually included was made by the senior psychiatrist on the basis of "clinical experience [p. 60]." Thus, in the absence of systematic sampling of items, no argument can be made for the content validity of the Midtown measure of psychological disorder. The same is true of the Stirling and Yoruba studies, where items were taken from the NSA and other test sources without explicit specification of the selection procedures (e.g., A. H. Leighton et al., 1963, p. 85; D. C. Leighton et al., 1963, pp. 202, 205).

It is doubtful whether content validity, in the strictest sense, can be achieved in the measurement of untreated psychological disorder, since there appears to be no universe of items which experts agree on as defining the variable. Four different sources have been cited in recent studies by the relatively few researchers who related their procedures to an established diagnostic system: the Sjöbring system used at Lund University, Sweden (Essen-Möller, 1956); the system used in the Department of Psychiatry of the National Taiwan University Hospital, Taipeh (Lin, 1953); the World Health Organization *International Classification of Diseases* (Primrose, 1962); and the American Psychiatric Association *Diagnostic and Statistical Manual* (D. C. Leighton et al., 1963). As Clausen points out (1961, pp. 131–132), the last two differ markedly as a function of the greater emphasis placed by European psychiatrists on hypothetical constitutional determinants.

Criterion-Oriented Validity: Concurrent and Predictive

Of the two types of criterion-oriented validity, concurrent and predictive (Cronbach &

---

[3] A preliminary report of a study from Poland (Gnat, Henisz, & Sarapata, 1964) describes procedures which appear similar to those used in Midtown and Stirling County. In addition, the Polish investigators made an attempt to cross-check these procedures against clinical examinations of subsamples of respondents. Since a detailed description of its research operations has not yet been published, however, this study is not included in the following methodological analysis.

Meehl, 1955), there is no evidence in the field studies for the latter. Typically conducted at one point in time, these studies have thus far not tested their assessments of disorder against criteria of future psychiatric condition, admission to treatment, or social functioning. What evidence, then, is provided for concurrent validity?

In Midtown, (Srole et al., 1962) the NSA and MMPI items proposed by the behavioral scientists were tested in a study involving 139 diagnosed neurotic and remitted psychotic patients, and 72 persons judged well by a psychiatrist on the basis of a half-hour interview. The result was that "almost all the NSA and MMPI symptom questions emerged with validity confirmed [p. 42]." Twenty-two of the items included in the final questionnaire discriminated between the patient and well groups at the .01 level of significance (Langner, 1962), and the remainder of the items contributed by the behavioral scientists, at the .05 level.[4] While the Midtown psychiatrists reported that, in rating cases well or not well, they gave special weight to 8 of the 22 items which discriminated at the .01 level, they also paid special attention to 6 items which failed to discriminate at this level (p. 396). Thus, in the Midtown Study, while the data from which the psychiatrists worked had been tested for concurrent validity in the manner described above, this test did not guide the use of these data by the psychiatrists.

In the Stirling County study, an attempt to identify valid items was made by administering NSA questions and items from other tests to untreated community samples and to patients diagnosed as neurotic. The selection of items included in the survey interviews, however, was not wholly determined by the results of this study (D. C. Leighton et al., 1963, p. 205). As in the Midtown Study, the Stirling County psychiatrists did not make use of objective scores based on these items in the judgmental assessment of psychiatric disorder.

Before considering how the psychiatrists actually used the symptom items in the Midtown and Stirling County studies, questions must be raised about concurrent validity of the items themselves. In attempts to validate these items, the patient criterion groups were homogeneous with regard to type of disorder (e.g., all neurotics in a study by Macmillan in D. C. Leighton et al., 1963, Chapter VII), or unspecified as to diagnostic composition (e.g., Manis, Brawer, Hunt, & Kercher, 1963). Moreover, the stubborn problem of well controls has not been met head on, much less solved. For example, in Langner's (1962) study identifying 22 items which discriminated at the .01 level between patient and well groups, the fact that the well group was identified by means of clinical judgment brings the problem back to its origin, since the items can be no more valid than the psychiatrist's judgments against which they were tested.

Nor does the solution appear to lie in avoiding clinical judgments by using an unselected sample of the nonpatient population as the healthy criterion group. Reports of community rates up to 64% from the field studies themselves argue against such a procedure. An attempt by Manis et al. (1963) to cross validate the 22 Midtown study items, using samples from patient and nonpatient populations as criterion groups, both illustrates the problem and raises another. Finding that a group of predischarge ward patients had an average symptom score lower than those of a community cross-section, and a group of college students, the authors argued that the result indicated a failure of the test since "there is little reason to believe that the mental health of these predischarge patients is equal to or better than the nonhospitalized populations [p. 111]." In the absence of independent evidence concerning the mental health of their nonpatient populations, however, it seems difficult to interpret this result. It is conceivable, though unlikely, that the predischarge ward patients are cured, while the nonpatients need treatment.

It is also possible, and perhaps more plausible, that the predischarge ward patients in the interest of "getting out" are simply less willing than nonpatients to admit socially undesirable behavior. If so, we are seeing evidence here of the impact of response style (cf. Jackson & Messick, 1962). What are the implications of such influence for the psychiatric evaluations which have relied heavily on the "face validity" of the items?

---

[4] Langner, Thomas S. Personal communication. February 1964.

The measure of disorder in the Midtown Study consisted of psychiatrists' ratings of the symptom data which ranged respondents on a scale from well through five degrees of severity of symptomatology: "mild," "moderate," "marked," "severe," and "incapacitated." Almost a quarter (23.4%) of the respondents were classified in the last three categories: marked, severe, and incapacitated. These are referred to collectively as "impaired," and are the "cases" in the Midtown Study. Michael, one of the evaluating psychiatrists on the Midtown Study, put it this way:

> The individuals in the Impaired category of mental health . . . are represented as being analogous to patients in psychiatric therapy. . . . When it is urged that the mental ratings "Marked" and "Severe" are comparable to the clinical conditions of patients in ambulatory treatment, and the rating "Incapacitated" to the clinically hospitalized, the distinction is presented . . . as an attempt to anchor our conceptualizations in relation to known degrees of psychopathology [Srole et al., 1962, p. 333].

There is evidence that this claim requires scrutiny. In the Midtown sample, 40 respondents reported being current outpatients in psychotherapy at the time of the interview; 182 reported that they were expatients. The evaluating psychiatrists had full knowledge of these facts when they made their judgments. Since expatients might be expected to have benefited from treatment, it is not remarkable to find that 54% of the 182 expatients in the Midtown sample were judged unimpaired. However, if respondents placed in the impaired categories indeed resemble psychiatric patients, as the Midtown researchers claim, it is hard to understand why the study psychiatrists placed 48% of the 40 *current* patients in the unimpaired categories (Srole et al., 1962, p. 147).

In the Stirling County study, disorder was defined in terms of judged similarity to descriptions in the 1952 *Diagnostic and Statistical Manual* of the American Psychiatric Association, rather than judged similarity to actual patients with whom the psychiatrists had had experience. The main rating was a psychiatric evaluation of "caseness" based on written summaries of symptom data collected for the most part by lay interviewers. It is described by D. C. Leighton et al. (1963) as "a rating of the probability that, at some time in his adult life, up to the time of the interview, the individual would qualify as a psychiatric case [p. 53]." The evidence for the validity of their conclusion "that at least half of the adults in Stirling County are *currently* suffering from some psychiatric disorder defined in the APA *Diagnostic and Statistical Manual* [p. 356]" rests largely on the study psychiatrists' blind evaluation of the likelihood that 47 former clinic patients, mostly neurotic, are cases. Of these 47, 81% were rated "almost certainly psychiatric," and 11% more were rated "probably psychiatric [p. 175]." Their problems, moreover, were viewed mainly as present rather than past (pp. 178–179). Thus the Stirling evaluators saw more disorder in their expatients than the Midtown evaluators saw in their current patients.

These results of the application of Midtown and Stirling County study evaluation procedures to patients and expatients suggest a number of possible interpretations: that patients get and remain sicker in Stirling County than in Midtown Manhattan; or that treatment in Stirling is less effective than in Midtown; or that Stirling methods are less able to distinguish between past and current problems; or that the definitions of cases are vastly different in the two studies; or that some combination of these circumstances has operated simultaneously. In brief, there is considerable ambiguity about the relations between the Midtown and Stirling judgmental ratings of untreated disorder.

There is, then, much to criticize and improve upon in these past attempts to investigate the criterion-oriented validity of both objective and judgmental measures of untreated disorder. There is, first of all, the absence of evidence of predictive validity, evidence that can only be supplied by prospective studies. Criterion-oriented attempts to establish both concurrent and predictive validity, however, face a common problem. Even with more attention, for example, to larger and diagnostically more heterogeneous patient criterion groups, independent criteria of "wellness," and problems of response style, there are strong reasons not to rely primarily on attempts to establish criterion-oriented validity. Foremost is the present fact that there are no generally agreed upon criteria of psychological health or disorder (cf. A. H. Leighton et al., 1963, p. 264).

Construct Validity

Cronbach and Meehl (1955) have argued that when no generally accepted criteria for the variable of interest are available, and when no universe of content is fully agreed upon as defining the variable, we must become interested in construct validity. These are the circumstances of untreated psychological disorder.

In Cronbach and Meehl's formulation,

> A necessary condition for a construct to be scientifically admissible is that it occur in a nomological net, at least *some* of whose laws involve observables [p. 290].

Furthermore:

> *unless the network makes contact with observations, and exhibits explicit, public steps of inference, construct validation cannot be claimed* [p. 291].

As noted above, the Midtown Study measure of mental health consisted of psychiatrists' ratings which ranged subjects on a scale from "well" through five degrees of severity of symptomatology: "mild," "moderate," "marked," "severe," and "incapacitated." The two rating psychiatrists (Srole et al., 1962) explained:

> Throughout the volumes of this Study, the data must be evaluated as *a rating of mental health based on the rating psychiatrists' perceptions operating through a questionnaire instrument* [p. 66].

Although they reported that positive responses to any of 14 specific items ordinarily precluded the classification of a subject as well, and positive responses to other items suggested various degrees of severity of symptomatology (pp. 396–397), the psychiatrists summarized their impression of the rating process:

> we used our clinical judgment to the best of our ability. It would be a mistake, however, to overlook the fact that their remain some aspects of the process which are not altogether in our awareness [pp. 62–63].

To the extent that this measurement of psychological disorder is private and hence not replicable, a claim for construct validity is precluded.

The same problem exists in the Leightons' Stirling County and Yoruba studies (1963). Although the problem of achieving public steps of inference in psychiatric evaluation concerned them, they did not attain this goal in either of these studies. Optimistically, they suggested that the development of the procedures to date:

> brings within sight the possibility that the evaluations could be done by a computer. To achieve this the steps would have to be broken down into even more specific items, and the intuitive leaps that are still allowed would have to be dissected so that their components could be identified [p. 267].

A computer program would certainly be a step toward construct validation of measures of psychological disorder. However, the construct validity of such a program could be evaluated only in relation to a nomological net. It is not clear in the work of the Leightons how such a net would be formulated. Although the Stirling County study was introduced with a theoretical volume (A. H. Leighton, 1959), the propositions in this theory are developed at a level of abstraction such that it does not make direct contact with observations. Instead, the guide for psychiatric evaluation was the American Psychiatric Association *Diagnostic and Statistical Manual* (1952). While this Manual often includes etiological propositions in its descriptions of nosological types, etiological inferences were avoided in the Stirling and Yoruba studies in the interest of interrater reliability. Thus, the psychiatric evaluations were removed entirely from a nomological framework within which construct validity could be evaluated.

Where, then, can we look for propositions placing psychological disorder in a nomological net? Not, it appears, to clinical experience. Unlike tuberculosis and pellagra, which are commonly cited as subjects of successful epidemiological research, psychological disorder does not constitute an etiologically defined disease entity (Clausen, 1961; Gruenberg, 1955). Instead:

> With symptoms still our primary basis for classification, we are at the same stage of knowledge about

mental disease that medicine occupied a century ago with reference to the "fevers." Typhoid, malaria, and a number of other diseases, all readily distinguishable now, were lumped together [Clausen, 1961, pp. 131–132].

Recognition of this situation has led a number of investigators to avoid psychodynamic inferences and assumptions in attempting to identify untreated disorder (e.g., A. H. Leighton et al., 1963, p. 89; D. C. Leighton et al., 1963, p. 48; New York State Department of Mental Hygiene, 1959, p. 83; Srole, et al., 1962, pp. 63, 134). In the absence of connections with individual psychodynamics, however, the meaning of the symptoms must emerge from the discovery of their relationships with sociocultural factors. It seems necessary, therefore, to look to the results of existing field studies for leads to a nomological network which can be used to validate the construct of psychological disorder.

RELATIONS BETWEEN SOCIOCULTURAL FACTORS AND SYMPTOMATOLOGY

Four sociocultural factors have been studied sufficiently frequently to show a pattern of

TABLE 3

MINIMUM AND MAXIMUM RATES OF PSYCHOLOGICAL DISORDER REPORTED ACCORDING TO AGE

	Minimum		Maximum		% dif.	Author(s)
	Age	%	Age	%		
	Maximum in adolescence					
Minimum in youngest group studied	0–5	0.03	11–15	1.88	1.85	Brugger, 1931
	0–4	0.4	10–14	4.0	3.6	Cohen et al., 1939
	0–5	0.0	11–15	7.5[a]	7.5	Brugger, 1937
	0–4	2.6	10–14	11.4	8.8	Roth & Luton, 1943
	0–5	0.2[a]	11–15	14.1[a]	13.9	Brugger, 1933
	Maximum in middle years					
	0–10	0.01	41–50	2.0[a]	2.0	Uchimara et al., 1940
	0–10	1.6[a]	61–70	7.2[a]	5.6	Tsuwaga et al., 1942
	0–14	0.8	15–34	14.8	14.0	Pasamanick et al.[b], 1959
	10-19[c]	13.9	40–49 & 50–59	28.1	14.2	Bremer, 1951
	15–24[c]	35.1[a]	25–34	83.3[a]	48.2	Llewellyn-Thomas, 1960
	Maximum in oldest group studied					
	20–29[c]	15.3	50–59[d]	30.8	15.5	Srole et al., 1962
	Under 39	46.3[a]	60+	63.7[a]	17.4	D. Leighton et al.[e], 1963
	16–24[c]	27.5[a]	65+	48.1[a]	20.6	Taylor & Chave[f], 1964
	18[c]	3.1	59[d]	26.4	23.3	Gnat et al., 1964
	0–4	0.9	80+	49.1	48.2	Essen-Möller, 1956
	Maximum in middle years					
Minimum in other than youngest group studied	60+	38.1[a]	40–59	40.5[a]	2.4	A. Leighton et al., 1963
	60–69	2.2[a]	70–79	5.4[a]	3.2	Akimoto et al., 1942
	40–44	8.7[a]	50–54	27.4[a]	18.7	Primrose, 1962

[a] Calculated by B. S. Dohrenwend.
[b] Excludes "other mental, psychoneurotic, and personality disorders" included in total rates.
[c] Youngest group reported.
[d] Oldest group reported.
[e] Results reported in A. Leighton et al., 1963, p. 152.
[f] Based on survey data without supplementary physicians' reports included in total rate.

TABLE 4

RATES OF PSYCHOLOGICAL DISORDER REPORTED FOR MALES AND FEMALES

Rate for males higher				Rate for females higher			
Percentage			Author(s)	Percentage			Author(s)
Male	Female	d		Male	Female	d	
1.1[a]	1.0[a]	0.1	Kaila, Study II	1.89[a]	1.91[a]	0.02	Rosanoff, 1917
1.3[a]	1.2[a]	0.1	Kaila, Study I	3.0[a]	3.1[a]	0.1	Tsuwaga et al., 1942
11.9	11.8	0.1	Fremming, 1951	1.5[a]	2.0[a]	0.5	Lemkau et al., 1942
1.4[a]	1.2[a]	0.2	Brugger, 1931	2.0[a]	2.7[a]	0.7	Eaton & Weil[b], 1955
1.0[a]	0.6[a]	0.4	Uchimara et al., 1940	22.1	24.0	1.9	Bellin & Hardt, 1958
2.8	1.9	0.9	Cohen et al., 1939	12.3	14.9	2.6	Essen-Möller, 1956
3.4[a]	2.0[a]	1.4	Akimoto et al., 1942	20.6	26.0	5.4	Bremer, 1951
7.9	6.0	1.9	Roth & Luton, 1943	10.0[a]	16.2[a]	6.2	Primrose, 1962
4.5[a]	2.5[a]	2.0	Brugger, 1937	6.0	12.3	6.3	Pasamanick et al., 1959
41.8[a]	38.9[a]	2.9	A. Leighton et al., 1963	47.0	65.0	18.0	D. Leighton et al.[c], 1963
9.0[a]	6.0[a]	3.0	Brugger, 1933	22.0	43.0	21.0	Taylor & Chave[d], 1964
				51.8[a]	75.2[a]	23.4	Llewellyn-Thomas, 1960

[a] Calculated by B. S. Dohrenwend.

[b] Rates for both sexes are higher than the overall rate in Table 1 because the overall rate is for prevalence in the summer of 1951, whereas sex rates had to be calculated from data on lifetime morbidity.

[c] Results reported in A. Leighton et al., 1963, p. 149.

[d] Based on survey data without supplementary physicians' reports included in total rate.

relationship to rates of judged psychopathology. These are age, sex, race, and socioeconomic status.

The age groups for which minimum and maximum rates have been reported are shown in Table 3. With only three exceptions, the minimum rate was in the youngest age group reported. However, Table 3 does not show a consistent pattern for the age at which maximum rates were found. In five of the studies, the maximum appeared in adolescence, in eight in the middle years, and in five in the oldest age group studied. No clues to etiology seem evident in such discrepant results.

Sex comparisons, shown in Table 4, likewise do not present a clear picture. The studies divide almost evenly between those which reported higher rates for men and those which reported higher rates for women.

The same absence of trends appears in the studies comparing rates for whites and Negroes. Of the eight studies in Table 5, four reported higher rates for whites and four for Negroes.

Against this background of inconsistent results, it is almost startling to find, as Table 6 shows, that 14 of the 18 studies which present data on the relationship with social class yielded the highest rate of judged psychopathology in the lowest economic stratum. Of the remaining 4, 2 of the early studies by Brugger, and Llewellyn-Thomas' Canadian

TABLE 5

RATES OF PSYCHOLOGICAL DISORDER REPORTED FOR WHITES AND NEGROES

Rate for whites higher				Rate for Negroes higher			
Percentage			Author(s)	Percentage			Author(s)
White	Negro	*d*		White	Negro	*d*	
1.9[a]	1.2[a]	0.7	Lemkau et al., 1942	2.2[a]	2.8[a]	0.6	Cohen et al., 1939
7.1	5.0	2.1	Rowntree et al., 1945	1.8[a]	7.0[a]	5.2	Rosanoff, 1917
7.8	4.2	3.6	Roth & Luton, 1943	11.1[a]	37.2	26.1	Hyde & Chisholm, 1944
11.2	4.6	6.6	Pasamanick et al., 1959	—	—	—	D. Leighton et al.[b], 1963

[a] Calculated by B. S. Dohrenwend.

[b] Results reported in ridits rather than percentages.

TABLE 6

MINIMUM AND MAXIMUM RATES OF PSYCHOLOGICAL DISORDER REPORTED ACCORDING TO SOCIOECONOMIC STATUS

	Percentage		d	Author(s)
	Minimum	Maximum		
	Minimum in highest stratum			
Maximum in lowest stratum	22.7[a]	23.6[a]	0.9	Bremer[b], 1951
	1.3[a]	3.2[a]	1.9	Akimoto et al.[c], 1942
	0.7	3.7	3.0	Cohen et al.[d], 1939
	1.1	6.6	5.5	
	30.0	37.8	7.8	Taylor & Chave[e], 1964
	7.3	16.6	9.3	Hyde & Kingsley, 1944
	17.4[a]	29.4[a]	12.0	Bellin & Hardt[b], 1958
	1.6	15.1	13.5	Gnat et al.,[f] 1964
	6.0	25.4	19.4	
	12.5	47.3	34.8	Srole et al., 1962
	—	—	—	Cole et al.[g], 1957
	—	—	—	D. Leighton et al.[h], 1963
	Minimum in a middle stratum			
	2.3[a]	2.9[a]	0.6	Tsuwaga et al.[c], 1942
	0.8	1.9	1.1	Lin, 1953
	1.2[a]	2.5[a]	1.3	Brugger[i], 1931
	10.3	29.7	19.4	Primrose, 1962
	Minimum in lowest stratum			
Maximum in a middle stratum	3.4[a]	5.3[a]	1.9	Brugger[i], 1937
	45.0	54.1[a]	9.1	Llewellyn-Thomas[j], 1960
	7.4[a]	25.7[a]	18.3	Brugger[i], 1933
	Minimum in a middle stratum			
Maximum in highest stratum	6.2	13.6	7.4	Pasamanick et al., 1959

[a] Calculated by B. S. Dohrenwend.

[b] Distribution of population in socioeconomic strata was reported only by number of families; since Japanese census reports do not include information on family size by SES, rates were calculated on the assumption of equal family size in all four strata. Rates for socioeconomic strata are below total rate for Tsuwaga due to reduction by 12 of number of cases reported according to SES.

[c] Subjects divided into only two strata.

[d] Data for whites only reported for two wards separately.

[e] Cole et al. do not report rates but state: "Four-fifths of the families in the lower social strata contained at least one mentally ill member, while less than one-half of the upper-stratum families were thus affected [p. 395]."

[f] Results reported in ridits rather than percentages.

[g] Occupations grouped by B. S. Dohrenwend into three strata: high (self-employed merchants, manufacturers, and farmers, and middle level civil servants), middle (merchants, manufacturers, and farmers employed by others, and low level civil servants), and low (workers and servants).

[h] Occupations grouped by B. S. Dohrenwend into three strata: high—independent business and salaried workers, middle—fisherman and farmers, low—laborers. The maximum rate in these strata is below the overall rate because of the exclusion of two categories of persons with high rates, grouped separately by Llewellyn-Thomas: i.e., housewives, with a rate of 76%, and miscellaneous, with a rate of 65%.

[i] Based on survey data without supplementary physicians' reports included in total rate.

[j] Rates for two cities reported separately.

village study, gave the highest rate in a middle stratum. The index of social class is somewhat problematical in these studies, however, since it is based on our grouping of the occupations reported by the authors, which may not reflect accurately the stratification in these rural areas. Only one study reported the highest rate in the highest income group.

It seems, therefore, in the face of other inconsistencies in the studies here reviewed, that a relationship of such apparent strength must command attention. The cumulative evidence it represents appears to establish the association of low socioeconomic status with a high rate of judged psychopathology as an important source of working hypotheses.

A THEORETICAL ISSUE

A number of investigators have pointed out that the relationship between social class and psychological symptoms can be explained either as social causation, with low status producing psychopathology, or as social selection, with preexisting psychological disorder determining social status (e.g., Dunham, 1961). The latter is, of course, compatible with the position that genetic factors are important in the etiology of psychological disorder.

In a notable advance over previous approaches, Srole and Langner in the Midtown Study (Srole et al., 1962, pp. 212–213, pp. 228–229) attempted to choose between these two interpretations by investigating parental class position, a factor clearly antecedent to the current psychiatric condition of their adult subjects. Finding a significant inverse relation between their subjects' symptomatology and the socioeconomic status of their subjects' parents, the investigators suggested that environmental deprivation in childhood is a causal factor in psychological disorder. However, they also found that the relation between parental socioeconomic status and impairing symptoms was weaker than the relation of own socioeconomic status to impairing symptoms. Moreover, subjects rated impaired were most likely to be found among people who were downwardly mobile relative to their parents, and least likely to be found among those who were upwardly mobile. Accordingly, the Midtown researchers suggested that perhaps both social causation, in the form of childhood deprivation, and social selection, in the form of intergeneration mobility, contribute to the strong inverse relation between rates of impairment and subjects' own socioeconomic status.

There are problems, however, in addition to lack of parsimony, with this explanation. Genetic predisposition could, with equal plausibility, be substituted in it for childhood deprivation. Moreover, symptoms judged psychotic appear to be an important exception. Langner, on the basis of further analysis concluded that:

> it is the adult life conditions in particular that stimulate the development of high rates of psychosis in the lower class; for the childhood conditions, at least as reported by our respondents, do not vary substantially between the classes [Langner & Michael, 1963, p. 454].

Implicit in Srole's and Langner's explanations is the assumption that impairing symptoms indicate a stable, self-perpetuating defect in the personalities of adult subjects. But what if we consider the possibility that many of these symptoms represent transient responses by normal personalities to objective, stress producing events, i.e. stressors (Dohrenwend, 1961) in the immediate environment (cf. Tyhurst, 1957, p. 161)? This hypothesis would provide a more parsimonious explanation of the array of Midtown findings concerning social status and psychological disorder. First, it suggests the possibility that the excess of cases in the lowest socioeconomic stratum is made up of normal persons reacting to unusually harsh and numerous stressors in their environment. Further, the increase in symptoms with downward mobility, and the decrease with upward mobility would be explained as consequences of changes in the number of harshness of stressors in the mobile subjects' environment. Hence, intergeneration mobility would be expected to produce a lower correlation between parental socioeconomic status and rate of impairing symptoms than between own socioeconomic status and rate of symptoms, the pattern of correlation that was found.

While this explanation in terms of contemporary stressors is parsimonious, it runs counter to the usual interpretation of symptomatology. For instance, in the Stirling County study, the distribution of cases according to age was interpreted as indicating that

> these symptoms in nonpatients have the same quality of persistence that has long been recognized as an outstanding feature of psychoneurosis when seen in clinics and private practice [D. C. Leighton et al., 1963, p. 358].

This inference was drawn, however, despite compelling evidence that some portion of the symptoms they observed were of the transient, situational variety. That is, in comparing the psychiatrists' evaluations based on interview data collected in 1952 with a general practitioner's independent diagnoses of 39 respondents about 4 years later, the authors noted that, of the 14 disagreements, 4 were due to "transient episodes of disorder (p. 196)." A fifth may also have been transient since he

> either had a better relationship to the general practitioner than to his predecessor or had actually improved, so that he changed from ill to not ill by 1958 [p. 197].

Moreover, there are strong, albeit indirect, empirical grounds for considering the possibility that the high level of symptomatology in the lowest socioeconomic stratum represents, at least in part, just such transient responses to stressors. These grounds are, first, evidence suggesting that stressors for which the individual cannot be held responsible are relatively pervasive in the lower class environment; second, observations that normal individuals exposed to stressors respond with symptoms which would ordinarily be classified as indicative of psychopathology; third, evidence of spontaneous recovery from these symptoms on removal of the stressors. Consider each in turn.

Relative Pervasiveness in Lower Class Environment of Stressors for which Individuals Are Not Responsible

Investigators of the quality of lower class life have not generally been able to disentangle the causal web of socioeconomic disadvantage and individual ineffectiveness (e.g., Knupfer, 1947). There are, however, some facts of lower class life which seem more plausibly interpreted as stressors in the environment than as defects in the individual.

Premature deaths of children and young adults, for example, are disruptive factors in family and social life over which survivors have little control. The comparatively short life expectancy in lower class groups (Mayer & Hauser, 1950) suggests that such stressors are relatively frequent there.

Unemployment and the threat of unemployment, also suffered more frequently in the lowest stratum than in other strata (e.g., Hollingshead, 1947; Langner & Michael, 1963, pp. 386, 388), similarly appear more dependent on environmental circumstance than upon individual failure. Wide fluctuations in rates of unemployment both over time and from one place to another support this interpretation.

Moreover, some stressors are likely consequences of residence in low income areas. Such areas, for example, are the targets of bulldozers, with attendant forced relocation of low income families (e.g., Fried, 1963). And it is into low income areas that new immigrant groups must frequently move, often giving rise to problems of conflict between members of different cultures (e.g., Dohrenwend & Smith, 1962).

Responses of Normal Persons to Stressors

Reid (1961) has reviewed the literature on the relation of immediate external events to the onset of mental disorders. He found that

> the epidemiological or statistical evidence about the effect of such externals on mental disease is relatively scanty . . . much of it comes from studies done in the war [p. 197].

Striking evidence that normal individuals respond to stressors with symptoms which are found in psychiatric patients is provided by Hastings' (1944) World War II study of 150 men in the Eighth Air Force who had completed their 30 mission tour of duty without reporting sick. Since men who are predisposed to breakdown were found in another study generally to report sick before flying half this many missions (Reid, 1961, p. 202), Hastings' finding that 95% of his subjects had developed symptoms of operational fatigue is particularly compelling. Hastings identifies operational fatigue as a condition which, unlike flying fatigue, is not cured by a few days rest, but can often be cured by a week and a half of therapy. It is a form of psychopathology attributed, in these cases, to the wartime flying experience rather than to previous neurotic symptoms or to family history.

Observations were also made during World War II of the responses of civilian populations to bombing. In a review of these materials, Janis concluded that repeated reports in both England and Germany indicate:

> Under conditions of severe bombing there is marked incidence of temporary emotional shock, presumably even among persons who were previously emotionally stable. Such reactions may take the form of excessive anxiety symptoms or of mild depression and apathy [pp. 96–97].

The investigators whom Janis reviewed had no direct information about the psychological condition prior to the experience of bombing of the persons whose reactions they observed. Bremer's (1951) study of a northern Norwegian village overcomes this problem. As resident physician from January 1939 to August 1945, he surveyed the entire population of 1400. When the village suffered enemy occupation and air raids, he found that 22 persons developed "war neurosis," i.e., "acquired nervous states caused by the direct effects of war," in which:

> the main syndrome was fear with its accompanying somatic conditions: starting, tremor, palpitation, precardiac pains. Add hereto in the majority of cases: fatigue, insomnia, and uncharacteristic dedolations. In a few cases the depression is extremely predominant.
>
> In two of the women, the nervous reaction is more correctly characterized as psychotic: in the one the syndrome was one of religious-ecstatic exaltation of one or two weeks' standing; the other suffered from a depression marked by anxiety, lasting 3–4 weeks [p. 57].

There have also been some relevant studies of reactions to stressors other than those produced by war. Lindemann (1944, pp. 146–147) followed up the survivors of the Coconut Grove fire in Boston. On the basis of his observations of these cases and others who had lost a friend or relative, he emphasized the importance of the situation, and irrelevance of personal factors, as determinants of extreme grief reactions.

Nor are such reactions observed only in situations involving loss of loved persons. Fried (1963) found indications of grief and mourning, similar to those described by Lindemann, in the majority of a sample of 566 men and women who had been forced to relocate from a slum section of Boston to make way for urban renewal. Their reactions included:

> the feelings of painful loss, the continued longing, the general depressive tone, frequent symptoms of psychological or social or somatic distress, the active work required in adapting to the altered situation, the sense of helplessness, the occasional expressions of both direct and displaced anger, and tendencies to idealize the lost place [p. 15].

Finally, in a study of reactions to President Kennedy's death (Sheatsley & Feldman, 1964), 89% of a national sample reported that during the 4 days following the assassination they experienced one or more of 15 physical and emotional symptoms such as "Didn't feel like eating," "Had headaches," "Had an upset stomach," "Had trouble getting to sleep," and "Felt nervous and tense." Most of these 15 items are similar to ones on which Midtown and Stirling County psychiatrists based their judgments of untreated psychological disorder (cf. Langner, 1962; D. C. Leighton et al., 1963, pp. 440–441; Srole et al., 1962, pp. 388–389).

The Reversibility of Stressor-Induced Symptomatology

If the symptoms normal persons develop when exposed to stressors disappear spontaneously on removal of the stressors, then it is clear that these symptoms indicate situationally induced response rather than self-perpetuating pathology in the individual. The weight of the evidence is that spontaneous recovery is frequent.

Thus Bremer (1951, p. 57), following the civilian cases of "war neurosis" he found in the northern Norwegian village, reports: "In all cases the prognosis was good, the symptoms disappearing when the patients were removed from the danger zone. In those who once more experienced war actions, the symptoms generally recurred.

Ginzberg and his colleagues (1959a, p. 19) studied the postwar adjustment of 534 "ineffective soldiers" systematically sampled from the men inducted in the last 4 months of 1942. An ineffective soldier was defined as "any man whom the Army discharged prior to demobilization for reasons of psychoneurosis, psychosis, inaptitude, or traits of character which made him unsuitable for retention in military service.

Relying mainly on Army and Veterans Administration records, Ginzberg and his colleagues classified types of readjustment after discharge into three major patterns: early, delayed, and unsuccessful. Early readjustment meant at least adequate civilian performance within 2 years of discharge. Delayed adjustment took longer than 2 years. Unsuccessful readjustment was indicated by failure of performance during the last 2 years for which information was available for a given soldier. The modal group, 44%, were judged to have made an early readjustment. Only 11.1% were in the delayed readjustment category, and 19.3% were judged to have made unsuccessful readjustments. For the remaining 25.6% of the sample, no data were available concerning adjustment.

With regard to loss of a cherished person or home, Lindemann (1944) reported, in a follow-up of 13 bereaved victims of the fire in the Coconut Grove, that all but one were judged, after a series of psychiatric interviews, to have made a satisfactory adaptation in 4 to 6 weeks. And Fried (1963) found that the minority, 26%, of the women in his sample of persons relocated from a Boston slum reported that they still felt sad or depressed 2 years later, with another 20% reporting periods of 6 months to 2 years. Slightly smaller percentages were found among the sample of 316 men.

In the study by Sheatsley and Feldman (1964) of reactions to President Kennedy's assassination, interviews were done 5 to 9 days after the day of the assassination. Although 89% of the sample had reported experiencing physical and emotional symptoms during the first 4 days, only 50% reported that they still had at least one symptom at the time of the interview.

SOME STRANDS IN A NOMOLOGICAL NETWORK

In a recent review of clinical observations about psychological disorder together with results from laboratory investigations of frustration, traumatic avoidance learning, and experimental neurosis, Wilson (1963) found common themes:

> On the one hand, the foundation for pathology is laid by a progressive state of emotional arousal that finally reaches disastrous proportions. . . . Secondly, the constant feature of the behavioral symptoms is their stereotypy and repetitiveness. Once established, the symptoms are remarkably intractable to control by external reward or punishment. . . . These . . . characteristics of behavior pathology, anchored as they are in careful experimental work, furnish substantial corroboration for . . . similar features noted . . . in the clinical literature [pp. 143–144].

Wilson's conclusion implies that the temporary, stressor-induced symptoms described in the preceding section should not be included under the rubric psychological disorder. Yet, in the absence of information as to whether the symptoms persist in the face of changed circumstances, it appears that situationally specific responses to stressors cannot be distinguished by clinicians from responses which outlive the conditions that induced them.

Faced with the problem of lack of information on persistence of symptomatology, one approach has been to interpret the symptoms relative to the setting in which they appear. Noyes and Kolb (1958) do this, for example, in suggesting that symptoms of normal battle reaction

> would be considered abnormal in a civilian setting yet are not so incapacitating as to demand removal from combat. Unless extreme, such reactions must *under the circumstances* be considered within the range of normal [p. 533, italics added].

In distinguishing here between normal and abnormal, Noyes and Kolb describe a process of social evaluation which has some interesting points of convergence with the conception of psychological disorder advanced by Wilson. Noyes and Kolb's definition suggests that symptoms are judged normal, in the particular setting of combat, when they are seen as not unduly damaging either to the military performance of the individual or to others in his group. Moreover, these authors' definition implies that when symptoms are judged harmful in these terms, it is necessary to alter the soldier's behavior, initially by removing him from combat.

Generalizing this process, we propose that psychological disorder can be defined, in part, by social judgments that the individual's reactions to a stressor in a particular social setting are harmful to him or to others in his group, and should be altered (cf. Dohrenwend, 1962, 1963; Hollingshead & Redlich, 1958, pp. 171–190). Relating this process of evaluation to Wilson's analysis, we suggest that a further condition for identification of psychological disorder is the intractable persistence of these symptomatic reactions in the face of the attempts by social agents to alter them.

These conditions can be combined into the following definition of psychological disorder as one approach to developing a nomological net:

Stressor-induced symptomatic responses, judged by the individual and/or by other social agents to be harmful to the individual and/or to others with whom he stands in social relationship, indicate psychological disorder if the symptoms continue (a) after the stressor ceases to impinge on the individual, and (b) despite sanctions directed at the individual by social agents who judge the response maladaptive.

This definition articulates within a single conceptual framework evidence concerning transient symptomatic responses to situational stressors with evidence, from both clinic and laboratory, which points to intractability to control by reward or punishment as the critical characteristic of symptoms indicative of psychological disorder.

A step toward validation of the construct, psychological disorder, could be taken, we suggest by using this definition in tests of the proposition that the high rates of "disorder" found by field studies in the lowest stratum represent, in large part, situationally specific symptomatic reactions rather than psychological disorder as defined above. Support for this proposition would be provided by confirmation of such hypotheses as the following:

1. That a large portion of the symptomatology reported in field studies is induced by stressors in the contemporary situation.

2. That a large portion of such stressor induced symptomatology is transient rather than persistent.

3. That transience of symptomatology varies directly with the extent to which it is defined as maladaptive.

4. That the extent to which such symptomatology is defined as maladaptive varies inversely with social class.

5. That the incidence of stressors in the contemporary situation varies inversely with social class.

Tests of these hypotheses would, we believe, provide a base from which to expand the nomological net to include crucial issues of the relative impact of genetic factors, childhood experiences, and contemporary environmental conditions on the development and course of psychological disorder.

REFERENCES

AKIMOTO, H., SIMAZAKI, T., OKADA, K., & HANASIRO, S. Demographische und psychiatrische Untersuchung über abgegrenzte Kleinstadtgevölkerung. *Psychiatria et neurologia Japonica*, 1942, **47**, 351–374.

AMERICAN PSYCHIATRIC ASSOCIATION COMMITTEE ON NOMENCLATURE AND STATISTICS. *Mental disorders: Diagnostic and statistical manual.* Washington, D. C.: American Psychiatric Association Hospital Service, 1952.

BELLIN, S. S., & HARDT, R. H. Marital status and mental disorders among the aged. *American Sociological Review*, 1958, **23**, 155–162.

BLUM, R. H. Case identification in psychiatric epidemiology: Methods and problems. *Milbank Memorial Fund Quarterly*, 1962, **40**, 253–288.

BREMER, J. A social psychiatric investigation of a small community in northern Norway. *Acta Psychiatrica et Neurologica Scandinavica*, 1951, *Supplement*, **62**, 1–166.

BRUGGER, C. Versuch einer geisteskrankenzählung in thüringen. *Zeitschrift für die gesamte Neurologie Psychiatrie*, 1931, **133**, 352–390.

BRUGGER, C. Psychiatrische ergebnisse einer medizinischen, anthropologischen, und soziologischen bevölkerunguntersuchung. *Zeitschrift für die gesamte Neurologie Psychiatrie*, 1933, **146**, 489–524.

BRUGGER, C. Psychiatrische bestandaufnahme im gebiet eines medizinisch-anthropoligschen zensus in der Nähe von Rosenheim. *Zeitschrift für die gesamte Neurologie Psychiatrie*, 1937, **160**, 189–207.

CARTWRIGHT, ANN. The effect of obtaining information from different informants on a family morbidity inquiry. *Applied Statistics*, 1957, **6**, 18–25.

CLAUSEN, J. A. Mental disorders. In R. K. Merton & R. A. Nisbet (Eds.), *Contemporary social problems.* New York: Harcourt, Brace, & World, 1961. Pp. 127–180.

COHEN, B. M., FAIRBANK, RUTH, & GREENE, ELIZABETH. Statistical contributions from the Eastern Health District of Baltimore. III. Personality disorder in the Eastern Health District in 1933. *Human Biology*, 1939, **11**, 112–129.

Cole, N. J., Branch, C. H. H., & Orla, M. Mental illness. *A.M.A. Archives of Neurology and Psychiatry,* 1957, **77,** 393–398.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, **52,** 281–302.

Dahlstrom, W. G., & Welsh, G. S. *An MMPI handbook.* Minneapolis: Univer. Minnesota Press, 1960.

Dohrenwend, B. P. The social psychological nature of stress: A framework for causal inquiry. *Journal of Abnormal and Social Psychology,* 1961, **62,** 294–302.

Dohrenwend, B. P. Some aspects of the appraisal of abnormal behavior by leaders in an urban area. *American Psychologist,* 1962, **17,** 190–198.

Dohrenwend, B. P. Urban leadership and the appraisal of abnormal behavior. In L. J. Duhl (Ed.), *The urban condition.* New York: Basic Books, 1963. Pp. 259–266.

Dohrenwend, B. P., & Smith, R. J. Toward a theory of acculturation. *Southwestern Journal of Anthropology,* 1962, **18,** 30–39.

Dunham, H. W. Social structures and mental disorders: Competing hypotheses of explanation. In *Causes of mental disorders: A review of epidemiological knowledge, 1959.* New York: Milbank Memorial Fund, 1961. Pp. 227–265.

Eaton, J. W., & Weil, R. J. *Culture and mental disorders.* Glencoe, Ill.: Free Press, 1955.

Essen-Möller, E. Individual traits and morbidity in a Swedish rural population. *Acta Psychiatrica et Neurologica Scandinavica,* 1956, *Supplement,* **100,** 1–160.

Faris, R. E. L., & Dunham, H. W. *Mental disorders in urban areas: An ecological study of schizophrenia and other psychoses.* Chicago: Univer. Chicago Press, 1939.

Felix, R. H., & Bowers, R. V. Mental hygiene and socio-environmental factors. *Milbank Memorial Fund Quarterly,* 1948, **26,** 125–147.

Fremming, K. H. *The expectation of mental infirmity in a sample of the Danish population.* London: Eugenics Society, 1951.

Fried, M. Grieving for a lost home. In L. J. Duhl (Ed.), *The urban condition.* New York: Basic Books, 1963. Pp. 151–171.

Ginzberg, E., Anderson, J. K., Ginsburg, S. W., & Herma, J. L. *Patterns of performance.* New York: Columbia Univer. Press, 1959. (a)

Ginzberg, E., Anderson, J. K., Ginsburg, S. W., & Herma, J. L. *The lost divisions.* New York: Columbia Univer. Press, 1959. (b)

Gnat, T., Henisz, J., & Sarapata, A. A psychiatric-socio-statistical study of two Polish towns. Paper read at First International Congress of Social Psychiatry, London, August 1964.

Gruenberg, E. M. Problems of data collection and nomenclature. In C. H. H. Branch, E. G. Beier, R. H. Anderson, & Carroll A. Whitmer (Eds.), *The epidemiology of mental health.* Brighton, Utah: Departments of Psychiatry and Psychology, University of Utah and Veterans Administration Hospital, Fort Douglas Division of Salt Lake City, Utah, 1955. Pp. 63–70.

Gurin, G., Veroff, J., & Feld, Sheila. *Americans view their mental health.* New York: Basic Books, 1960.

Hastings, D. W. Psychiatry in Eighth Air Force. *Air Surgeon's Bulletin,* 1944, **1**(8), 4–5.

Hollingshead, A. B. Selected characteristics of classes in a middle western community. *American Sociological Review,* 1947, **12,** 385–395.

Hollingshead, A. B., & Redlich, F. C. *Social class and mental illness.* New York: Wiley, 1958.

Hyde, R. W., & Chisholm, R. M. The relation of mental disorders to race and nationality. *New England Journal of Medicine,* 1944, **231,** 612–618.

Hyde, R. W., & Kingsley, L. V. Studies in medical sociology: The relation of mental disorder to the community socioeconomic level. *New England Journal of Medicine,* 1944, **231,** 543–548.

Jackson, D. N., & Messick, S. Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology,* 1962, **65,** 285–299.

Janis, I. L. *Air war and emotional stress.* New York: McGraw-Hill, 1951.

Kaila, M. Über die Durchschnittshäufigkeit der Geisteskrankheiten und des Schwachsinns in Finnland. *Acta Psychiatrica et Neurologica,* 1942, **17,** 47–67.

Knupfer, Genevieve. Portrait of the underdog. *Public Opinion Quarterly,* 1947, **11,** 103–114.

Langner, T. S. A twenty-two item screening score of psychiatric symptoms indicating impairment. *Journal of Health and Human Behavior,* 1962, **3,** 269–276.

Langner, T. S. Psychophysiological symptoms and women's status in Mexico. In Jane M. Murphy & A. H. Leighton (Eds.), *Approaches to cross-cultural psychiatry,* in press.

Langner, T. S., & Michael, S. T. *Life stress and mental health.* New York: Glencoe, Ill.: Free Press, 1963.

Leighton, A. H. *My name is legion.* New York: Basic Books, 1959.

Leighton, A. H., Lambo, T. A., Hughes, C. C., Leighton, Dorothea C., Murphy, Jane M., & Macklin, D. B. *Psychiatric disorder among the Yoruba.* Ithaca, N. Y.: Cornell Univer. Press, 1963.

Leighton, Dorothea C., Harding, J. S., Macklin, D. B., Macmillan, A. M. & Leighton, A. H. *The character of danger.* New York: Basic Books, 1963.

Lemkau, P., Tietze, C., & Cooper, M. Mental hygiene problems in an urban district. *Mental Hygiene,* 1941, **25,** 624–646.

Lemkau, P., Tietze, C., & Cooper, M. Mental-hygiene problems in an urban district: Second paper. *Mental Hygiene,* 1942, **26,** 100–119.

Lin, Tsung-yi. A study of the incidence of mental disorder in Chinese and other cultures. *Psychiatry,* 1953, **16,** 313–336.

Lindemann, E. Symptomatology and management of acute grief. *American Journal of Psychiatry,* 1944, **101,** 141–148.

Llewellyn-Thomas, E. The prevalence of psychiatric symptoms within an island fishing village. *Canadian Medical Association Journal,* 1960, **83,** 197–204.

Manis, J. G., Brawer, M. J., Hunt, C. L., & Kercher, L. C. Validating a mental health scale. *American Sociological Review,* 1963, **28,** 108–116.

Manis, J. G., Brawer, M. J., Hunt, C. L., & Kercher, L. C. Estimating the prevalence of mental illness. *American Sociological Review,* 1964, **29,** 84–89.

Mayer, A. J., & Hauser, P. Class differentials in expectation of life at birth. *Review. Institute of International Statistics,* 1950, **18,** 197–200.

Mayer-Gross, W. Mental health survey in a rural area. *Eugenics Review,* 1948, **40,** 140–148.

New York State Department of Mental Hygiene, Mental Health Research Unit. A mental health survey of older people. *Psychiatric Quarterly Supplement,* 1959, Part I, 45–99.

Noyes, A. P., & Kolb, L. C. *Modern clinical psychiatry.* (5th ed.) Philadelphia: Saunders, 1958.

Pasamanick, B. Thoughts on some epidemiologic studies of tomorrow. In P. H. Hoch & J. Zubin (Eds.), *The future of psychiatry.* New York: Grune & Stratton, 1962. Pp. 216–223.

Pasamanick, B., Roberts, D. W., Lemkau, P. W., & Krueger, D. B. A survey of mental disease in an urban population: Prevalence by race and income. In B. Pasamanick (Ed.), *Epidemiology of mental disorder.* Washington, D. C.: American Association for the Advancement of Science, 1959. Pp. 183–191.

Plunkett, R. J., & Gordon, J. E. *Epidemiology and mental illness.* New York: Basic Books, 1960.

Primrose, E. J. R. *Psychological illness: A community study.* London: Tavistock Publications, 1962.

Reid, D. D. Precipitating proximal factors in the occurrence of mental disorders: Epidemiological evidence. In *Causes of mental disorders: A review of epidemiological knowledge, 1959.* New York: Milbank Memorial Fund, 1961. Pp. 197–216.

Rosanoff, A. J. Survey of mental disorders in Nassau County, New York, July–October 1916. *Psychiatric Bulletin,* 1917, **2,** 109–231.

Roth, W. F., & Luton, F. B. The mental hygiene program in Tennessee. *American Journal of Psychiatry,* 1943, **99,** 662–675.

Rowntree, L. G., McGill, K. H., & Hellman, L. P. Mental and personality disorders in selective service registrants. *Journal of the American Medical Association,* 1945, **128,** 1084–1087.

Sandifer, M. G., Jr. Social psychiatry 100 years ago. *American Journal of Psychiatry,* 1962, **118,** 749–750.

Sheatsley, P. B., & Feldman, J. The assassination of President Kennedy: Public reactions. *Public Opinion Quarterly,* 1964, **28,** 189–215.

Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., & Rennie, T. A. C. *Mental health in the metropolis: The midtown study,* Vol. I. New York: McGraw-Hill, 1962.

Star, Shirley A. The screening of psychoneurotics in the army: Technical development of tests. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, Shirley A. Star, & J. A. Clausen (Eds.), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 486–547. (a)

Star, Shirley A. The screening of psychoneurotics: Comparison of psychiatric diagnoses and test scores at all induction stations. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, Shirley A. Star, & J. A. Clausen (Eds.), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 548–567. (b)

Strömgren, E. Statistical and genetic population studies within psychiatry; methods and principal results. *Actualités scientifiques et industrielles, Congrès International de Psychiatrie, VI, Psychiatrie Sociale.* Paris: Herman & Cie., 1950. Pp. 155–188.

Szasz, T. S. *The myth of mental illness.* New York: Paul B. Hoeber, 1961.

Taylor, Lord, & Chave, S. *Mental health and environment.* London: Longmans, Green, 1964.

Trussell, R. E., Elinson, J., & Levin, M. L. Comparisons of various methods of estimating the prevalence of chronic disease in a community—The Hunterdon County Study. *American Journal of Public Health,* 1956, **46,** 173–182.

Tsuwaga, T., Okada, K., Hanasiro, S., Asai, T., Takuma, T., Morimura, S., & Tsuboi, F. Über die psychiatrische zensus-untersuchung in einem stadtbezirk von Tokyo. *Psychiatria et neurologia Japonica,* 1942, **46,** 204–218.

Tyhurst, J. S. The role of transition states—including disasters—in mental illness. In *Symposium on preventive and social psychiatry.* Washington, D. C.: United States Government Printing Office, 1957. Pp. 149–169.

Uchimara, Y., Akimoto, H., Kan, O., Abe, Y., Takahashi, K., Inose, T., Shimazaki, T., & Ogawa, N. Über die vergleichend psychiatrische und erbpathologische untersuchung auf einer Japanischen insel. *Psychiatria et neurologia Japonica,* 1940, **44,** 745–782.

Wilson, R. S. On behavior pathology. *Psychological Bulletin,* 1963, **60,** 130–146.

(Received March 19, 1964)

length of hospitalization for the group was 199.4 days. Their psychiatric diagnoses consisted of 10 schizophrenics, 4 psychotic depressions, and 2 character disorders. Again, control data indicated no significant differences between the two compatibility groups as regards age, length of hospitalization, distribution of psychiatric diagnoses, and administration of drugs.

Part C utilized the same 16 patients who appeared in the discharge-testing phase of Part B.

Doctors

The doctors consisted of one first and one second year male psychiatric resident, and a first year female psychiatric resident. Of the 25 patients, 8 had been randomly assigned to each of two residents, and 9 to the third. Each patient was routinely seen by her doctor every morning for a short time during rounds, and for at least 1½ hours per week in psychotherapy sessions.

Materials

Compatibility scores for patient-doctor combinations were obtained by means of the FIRO-B scale. This scale identifies three areas of interpersonal need (Inclusion, Control, Affection), claimed to constitute a sufficient set of areas of interpersonal behavior for the prediction and explanation of interpersonal phenomena (Schutz, 1958, p. 13). Raw scores were derived for each need area based upon how an individual stated she treated others and how she wished to be treated by them. The raw scores for each patient and her doctor were combined in the manner described by Schutz (p. 105), yielding a compatibility score for each of the three interpersonal need areas. The three compatibility scores were then combined to give a Total Compatibility score (K). All *K* scores used in the study resulted from the initial testing of both patient and doctor. The smaller the *K* score, the greater was the dyadic compatibility. These measures were utilized in Parts A and B of the study.

Parts B and C concerned themselves with how each patient perceived and felt perceived by her doctor. A Semantic Differential Scale was developed to obtain this data. The choice of concepts for this scale was based upon an expectation that a number of them would be of especial importance to each of the patients, and would constitute some of the content in their psychotherapy sessions. The scale consisted of 11 concepts. These were: Spouse, Child, Doctor, Anger, Death, Mother, Sex, Love, Patient, Me, and Father. Each concept was rated on seven-point scales consisting of items heavily weighted in each of three factors identified in Semantic Differential construction. Those items heavily weighted in the Evaluative factor were: fair-unfair, sour-sweet, and dirty-clean. Those weighted in the Potency factor were: soft-hard, strong-weak, and light-heavy. The Activity factor was represented by: hot-cold, slow-fast, and sharp-dull. These items were placed randomly under each of the 11 concepts.

Procedure

It was believed that a 1-month period would provide the patient sufficient time for an initial adjustment to her new surroundings and would enable both patient and doctor to have formed impressions of one another. After 4 weeks hospitalization, the patient was asked by the investigator to participate in a research project examining psychotherapy. Patients were assured that their communications would be held in strict confidence and would not be communicated to their doctors by the investigator, but that patients were free to discuss whatever transpired with their doctors if they wished. After a patient agreed to participate, the FIRO-B and Semantic Differential Scales were individually administered by the investigator.

The FIRO-B scale was administered to the doctors at the outset of the study. They were informed that this was a research effort to gain a better understanding of the patient-doctor interaction, and that they would be exposed to the same procedure as their patients. They were also told that the patients' data could not be made available to them, and that they should not volunteer the information to the patients that they too were participants in the study. The Semantic Differential was presented whenever each of their new patients had been in the hospital for 4 weeks.

Three instructional "sets" were given patients for completing the Semantic Differential. Though independently derived, they bear a similarity to Fiedler's (1951) procedure. The instructions were: (*a*) fill it out for yourself, (b) fill it out the way your doctor would fill it out for himself, and (*c*) fill it out in the way your doctor would fill it out for you. This yielded three combinations from which total difference scores (*D*) were obtained. These combinations were: (*a*) the difference between patient's ratings of self and patient's ratings of doctor (the smaller the *D*, the more the experienced similarity in relation to the rated concepts), (*b*) the difference between patient's ratings of doctor and patient's ratings of the way she thought the doctor saw her (the smaller the *D*, the more similar she felt the doctor experienced them to be in terms of the rated concepts), (*c*) the difference between patient's ratings of self and patient's ratings of the way she thought the doctor saw her (the smaller the *D*, the more understanding the patient felt the doctor had of her). A *D* score was obtained, for each combination, in each of the three Semantic Differential factors. Thus, a total of nine *D* scores was obtained for correlation with compatibility (*K*) scores in Part B and supervisors' ratings of improvement in Part C.

Comparable instructional sets were given to the doctors, for each of their patients, when they filled out the Semantic Differential. The resultant *D* scores were similarly treated.

The entire procedure was repeated 2 weeks prior to the scheduled discharge of each patient from the hospital.

After a patient had been discharged from the hospital, the senior psychiatrist, who had continued independent contact with the patient on the living unit as well as having supervised the treating resident, rated the degree of improvement shown by the patient. Ratings were done on an eight-point, self-anchored scale (Gardner & Thompson, 1956). There were two supervising psychiatrists, and each anchored their rating scale, at the outset of the investigation, with the names of former patients whose cases were very familiar to them. These "anchor" patients were chosen so as to represent points along a continuum of improvement. Thirteen patients were rated by one supervisor and nine by the other. A Mann-Whitney U test yielded no significant difference between supervisors' ratings, suggesting that they both tended to rate from a similar frame of reference.

The average discharge rating given to each of the residents' patients was 6.0, 6.0, and 5.7, respectively. This reflected a similarity in outcome of treatment for each resident, permitting the combining of their data for statistical analyses.

RESULTS

Part A

A Pearson product-moment correlation of .45 was obtained between patient-doctor compatibility scores (K) and supervisors' ratings of patient improvement. The r of .45, with 20 *df*, was significant beyond the .05 level of statistical confidence for a two-tailed test. This finding confirmed the hypothesis that compatibility of interpersonal needs would be positively related to the outcome of treatment. Highly compatible patients showed greater effects of their doctor's influence than did low compatibility patients.

To check whether the significant correlation arose through the compatibility interaction between patient and doctor and not as the result of some uncontrolled factor distinguishing doctors from patients, an additional analysis was made by assigning As patients to B, Bs to C, Cs to A. The new compatibility scores, derived from the combination of doctors and patients who had not worked together, were correlated with ratings of improvement. This procedure yielded an r of $-.28$, indicating that the obtained significant correlation did indeed result from what had transpired in the interrelationships between the patient and her own doctor.

Part B

It was hypothesized that the greater the degree of compatibility, the smaller would be the differences in Semantic Differential scores between the three instructional sets. The product-moment correlations obtained between patient-doctor compatibility scores (K) and Semantic Differential difference scores (D) appear in Table 1. At the time of *initial* testing, no combination, for either patient or doctor, on any of the three Semantic Differential factors, revealed correlations which were statistically significant. Thus, the compatibility factor was found to be unrelated to the manner in which patient or doctor perceived each other during the first month of hospitalization.

At the time of *discharge* testing, two of the three patient D scores in the Evaluative factor showed statistically significant correlations with K at the .05 level of confidence (14 *df*) for a two-tailed test. This confirmed the hypotheses that the more compatibility between patient and doctor, the more such a patient would feel that: (*a*) she was being understood by her doctor, and (*b*) a similarity existed between herself and her doctor.

All other patient and doctor D scores were not found to be significantly related to the compatibility variable.

To identify trends which were themselves suggestive but failed to attain necessarily rigorous statistical standards, a one-tailed test of significance was applied to the discharge data in Table 1. With 14 *df* an r of .42 was significant at the .05 level of confidence. Using this criterion, all three patient Evaluative combinations and one patient Potency combination were identified. This suggests that there may have been an effect of the independent variable which required a larger N before it became manifest statistically. No doctor combination evidenced a significant correlation.

As an additional evaluation of trend, a sign test was applied to the nine paired initial-discharge Semantic Differential combinations for both patients and doctors. The sign test for patient pairs was significant beyond the .01 level of confidence. This suggests that the compatibility variable was operative within all three Semantic Differential factors at the time of discharge. The sign test was insignificant for doctor pairs.

TABLE 1

PATIENT AND DOCTOR CORRELATIONS BETWEEN COMPATIBILITY SCORES AND DIFFERENCE SCORES OF SEMANTIC DIFFERENTIAL COMBINATIONS ARISING FROM THREE INSTRUCTIONAL "SETS"

Semantic differential combinations	Semantic differential factors					
	Evaluative		Potency		Activity	
	df	*r*	*df*	*r*	*df*	*r*
Initial testing						
Patients						
PS:PD[a]	23	−.03	23	−.15	23	−.15
PD:PGD	23	−.07	23	−.34	23	−.24
PS:PGD	23	.00	23	−.33	23	−.19
Doctors						
DS:DP	23	.12	23	.31	23	.13
DP:DGP	23	−.08	23	−.20	23	.11
DS:DGP	23	−.14	23	.30	23	.34
Discharge testing						
Patients						
PS:PD	14	.56*	14	.45	14	.30
PD:PGD	14	.47	14	.30	14	.38
PS:PGD	14	.49*	14	.02	14	.23
Doctors						
DS:DP	14	.40	14	−.04	14	.30
DP:DGP	14	−.14	14	.02	14	−.37
DS:DGP	14	−.12	14	.01	14	.26

[a] PS = Patient's self-rating; PD = Patient's rating of doctor for himself; PGD = Patient's guess of doctor's rating of her.
* $p \leq .05$.

Part C

It was hypothesized that the smaller the differences in Semantic Differential scores between the three instructional sets, the greater would be the degree of patient improvement upon discharge. The product-moment correlations obtained between *D* scores and supervisors' ratings of patient improvement appear in Table 2. One of the three Semantic Differential combinations for patients showed statistically significant correlations at the .05 level of confidence (14 *df*) for a two-tailed test. The significant correlations were obtained in the Evaluative and Activity factors for patient-self:patient-doctor ratings. This confirmed the hypothesis that greater im-

TABLE 2

PATIENT AND DOCTOR CORRELATIONS BETWEEN SUPERVISORS' IMPROVEMENT RATINGS AND DIFFERENCE SCORES OF SEMANTIC DIFFERENTIAL COMBINATIONS ARISING FROM THREE INSTRUCTIONAL "SETS"

Semantic differential combinations	Semantic differential factors					
	Evaluative		Potency		Activity	
	df	*r*	*df*	*r*	*df*	*r*
Patients						
PS:PD[a]	14	.53*	14	.38	14	.50*
PD:PGD	14	.46	14	.22	14	.44
PS:PGD	14	.42	14	.36	14	.35
Doctors						
DS:DP	14	.43	14	.15	14	.27
DP:DGP	14	.24	14	.00	14	.16
DS:DGP	14	−.18	14	−.21	14	−.26

[a] PS = Patient's self-rating; PD = Patient's rating of doctor for himself; PGD = Patient's guess of doctor's rating for her.
* $p \leq .05$.

provement would be observed in patients who experienced themselves as similar to their doctor.

All other patient and doctor *D* scores were not found to be significantly related to improvement.

To identify trends, a one-tailed test of significance was employed. All three patient combinations in the Evaluative factor and two of the three combinations in the Activity factor were identified.

One combination in the Evaluative factor for doctors was also identified; doctor-self: doctor-patient. This same combination also showed the highest correlation in the discharge testing in Table 1. Thus, it may have been that doctors did tend to view themselves as similar to their most improved patients.

DISCUSSION

The findings indicated that the degree of interpersonal compatibility existing between patient and doctor was a significant variable affecting the outcome of hospital treatment. This appears to have occurred because of the effect the compatibility factor had upon the way the resident was perceived by the patient. Highly compatible patients were more likely to feel that a similarity existed between themselves and their doctors and that their doctors understood them. Low compatibility dyads resulted in those patients feeling that a dissimilarity existed between themselves and their doctors and that their doctors did not understand them. As a group, high compatibility patients displayed more improvement in their condition than did low compatibility patients. Thus, despite the many other uncontrolled factors also affecting outcome of treatment, the compatibility variable stood out. Limitation of sample size makes conclusions tentative. Further substantiation of the findings is required before implications can be drawn for patient-doctor assignments. It would appear, however, that an important underlying personality variable contributing to the establishment of "good" therapeutic relationships (Fiedler, 1950; Parloff, 1961) may have been identified.

The change in patients' perception of their doctors was found to be an evolving process. Initially no differences were evident in the attitudes of high and low compatibility patients, but at the time of discharge a divergent effect had occurred. The process may be compared to a finding in the verbal conditioning study where an initial delay occurred before there was a divergence in the learning curves of compatible and incompatible subjects. This was interpreted as being due to the intrinsic quality of the compatibility variable, which required time in order for the emotional situation to crystallize itself and become evident in the subjects' attitude toward their experimenters. It would also appear that some form of incubation period is necessary before the directional process becomes manifest in the patient-doctor interaction.

The positive trend found in the correlation between improvement and patients' feeling understood by their doctors, offered some experimental support for a commonly held clinical assumption. Patients' feeling understood by their doctors was also found to be significantly related to the compatibility variable. It is interesting to speculate that this finding may not have been forthcoming from patients with less disturbed reality testing, who would perhaps require a more overt indication that they were being understood than would psychotic patients. On the other hand, since psychotics are often thought to have a heightened sensitivity to the reactions of others, there may actually have been some differential responding by doctors which was not tapped in the present data.

Although differences in doctors' perceptions of patients were not significantly related to the compatibility or improvement variables, it must be pointed out that some suggestion of a differential-attitudinal orientation was noted. The indication of such a trend occurred in the doctor-self: doctor-patient combination, which suggested that doctors may have felt that some similarity existed between themselves and their more improved patients. Such an occurrence would be in keeping with the findings of Fiedler and Senior (1952). A similar trend was not obtained in the two other doctor combinations, indicating that they did not feel these patients understood them or considered themselves as similar to the doctor. One can speculate that for persons relatively new to the field, too much felt similarity with patients was difficult to accept. Thus, it might be that more experienced practitioners would be freer in revealing felt similarities, if indeed they existed.

The sign test used to determine the trend in paired correlations from initial to discharge testing indicated that the compatibility variable was operative within all three patient Semantic Differential factors. Yet, it was clear from the data that the Evaluative factor was most sensitive to attitudinal change. Why this was so challenges explanation. Perhaps the emotional loading of this factor provides for the easiest expression of positive and negative attitudes. This would be in keeping with Osgood's (1957, p. 190) identification of attitude with the Evaluative dimension.

It should be noted that incompatible subjects in the verbal conditioning study showed evidence of having been influenced to the same extent as compatible subjects, but this was displayed only in the extinction phase when not in the presence of their experimenters. This can suggest that some of the low compatibility patients, who may not have shown the full effect of their doctors' influence while in the hospital, may nevertheless have been positively influenced. It could be that for such patients the influence experienced during hospitalization would be expressed only after termination of the therapeutic relationship.

REFERENCES

FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, non-directive and Adlerian therapy. *Journal of Consulting Psychology*, 1950, **14**, 436–445.

FIEDLER, F. E. A method of objective quantification of certain countertransference attitudes. *Journal of Clinical Psychology*, 1951, **7**, 101–107.

FIEDLER, F. E., & SENIOR, K. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. *Journal of Abnormal and Social Psychology*, 1952, **47**, 446–453.

FREUD, S. *An outline of psychoanalysis*. New York: Norton, 1949.

GARDNER, E. F., & THOMPSON, G. G. *Social relations and morale in small groups*. New York: Appleton-Century-Crofts, 1956.

MARMOR, J. A reevaluation of certain aspects of psychoanalytic theory and practice. In L. Salzman & J. H. Masserman (Eds.), *Modern concepts of psychoanalysis*. New York: Philosophical Library, 1962.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

PARLOFF, M. B. Therapist-patient relationships and outcome of psychotherapy. *Journal of Consulting Psychology*, 1961, **25**, 29–38.

ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.

SAPOLSKY, A. Effect of interpersonal relationships upon verbal conditioning. *Journal of Abnormal and Social Psychology*, 1960, **60**, 241–246.

SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.

SHOBEN, E. J. Psychotherapy as a problem in learning theory. *Psychological Bulletin*, 1949, **46**, 366–392.

STRUPP, H. H. Psychotherapy. In P. R. Farnsworth, O. McNemar, & Q. McNemar (Eds.), *Annual review of psychology*. Palo Alto, Calif.: Annual Reviews, 1962.

SULLIVAN, H. S. *Clinical studies in psychiatry*. New York: Norton, 1956.

(Received March 26, 1964)

Saunders Books

New from Saunders!

Rosen and Gregory—
Abnormal Psychology
A clear, dynamic presentation of psychological disturbances

This unusual new text provides the student with new insight into the main syndromes of abnormal behavior. The authors examine such vital questions as "What is abnormality?" . . . "What abnormal behavior is associated with a particular diagnosis?" . . . "What are the characteristics, dynamics, or modes of adaptation?" . . . "What factors initiated or perpetuated abnormal behavior?" . . . "What is the developmental history?" . . . "What treatment methods are available?"

The authors first present a clear picture of general principles of abnormal psychology. They then apply these principles to analyses of clinical signs and symptoms. A great many illustrative cases cover almost every type of abnormal behavior—particular problems such as: *Disorientation—Confabulation—Depression—Delusions of grandeur—Stuttering—Compulsion.* Professional jargon has been minimized, and technical terms are clearly defined as they are introduced.

This outstanding text will synthesize for your students the important psychological concepts essential to the study and appreciation of abnormal psychology.

By the late EPHRAIM ROSEN, Ph.D., Professor of Psychology, University of Minnesota; and IAN GREGORY, M.D., Associate Professor of Psychiatry, University of Minnesota. About 576 pages, 7¼" x 10¼", illustrated. About $8.00. *New—Just Ready!*

Reed and Reed—
Mental Retardation:
A Family Study
A thorough genetic tabulation

This fascinating volume reports on the incidence of mental retardation and psychiatric disorders through several generations in entire families—*the results of a continuous study begun over 50 years ago.* The monumental documentation contains comprehensive genetic and social information about some 80,000 persons, providing unparalleled depth in research and testing data for biologists, geneticists, psychologists, and sociologists.

More than 400 pedigree charts are reproduced from original studies on patients at a Minnesota State Hospital, including their grandparents, immediate families, and progeny.

Significant observations are drawn on amelioration of mental conditions throughout generations—on qualities of skill and leadership despite moronic intellect—on the low rate of reproduction among families—on involvement of transmission in the etiology of retardation—on relative influences of genetic and environmental factors.

Psychologists and sociologists will find definitive and extraordinarily detailed material on the course of natural selection as related to intelligence.

By ELIZABETH W. REED, Ph.D., and SHELDON C. REED, Ph.D. Both at Dight Institute for Human Genetics, University of Minnesota, Minneapolis, Minnesota. About 800 pages, 7¼" x 10¼", with about 290 figures. About $18.00 *New—Just Ready!*

Texts gladly sent to professors on approval

W. B. SAUNDERS COMPANY W. Washington Square, Phila., Pa. 19105

Discount accorded to full-time instructors listing affiliation

Please send and bill me:

☐ Rosen & Gregory—**Abnormal Psychology**—about $8.00
☐ Reed & Reed—**Mental Retardation**—about $18.00

Name________________ Position________________
Address________________ JASP2-65

The Journal of

conflict RESOLUTION

A quarterly for research related to war and peace

DECEMBER 1964 ISSUE (VIII, 4):

SOCIAL VALUES AND FOREIGN POLICY ATTITUDES OF SOVIET AND AMERICAN ELITES

Including content-analytic studies by ROBERT C. ANGELL and J. DAVID SINGER; study of Soviet literature by VERA S. DUNHAM.

Book reviews by Allen D. Grimshaw, Richard W. Leopold, James K. Pollock, J. Alan Winter. *$5.50 year, $2.00 for single issues.*

Published by The Center for Research on Conflict Resolution
The University of Michigan, Ann Arbor, Michigan 48104

BEHAVIOR DISORDERS: Perspectives and Trends

Edited by Ohmer Milton, University of Tennessee

The editor has selected twenty-six articles from journals and other sources in which the limitations of the disease view of behavior disorders are measured against more recent conceptual schemes that explore disorders from the larger perspectives of family and community. An introductory chapter examines the general problem of classification of disorders, and each selection is prefaced by a precis of its content. A stimulating supplement for courses in Abnormal Psychology that will adjust well to any standard text.

Paperbound

$3.50

THE LIPPINCOTT PSYCHOLOGY SERIES UNDER THE EDITORSHIP OF JULIUS WISHNER, UNIVERSITY OF PENNSYLVANIA, AND CARL P. DUNCAN, NORTHWESTERN UNIVERSITY

Write for examination copies

College Department

J. B. LIPPINCOTT COMPANY
East Washington Square
Philadelphia, Pa. 19105

Volume 70 April 1965 Number 2

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE BURDOCK
New York State Psychiatric Institute

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

NICHOLAS HOBBS
George Peabody College

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The Journal is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

Helen Orr
Managing Editor

Robert L. Hood
Circulation Manager

Elizabeth S. Reed
Advertising Manager

Frances L. Brewer
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

The American Psychological Association, Inc.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 77–106

PSYCHOLOGICAL DEFICIT IN SCHIZOPHRENIA:

II. INTERFERENCE AND ACTIVATION[1]

PETER J. LANG[2] AND ARNOLD H. BUSS

University of Pittsburgh

This is the 2nd of 2 papers reviewing laboratory studies of psychological deficit in schizophrenia. The present report first considers experiments on attention, set, and association. Research based on drive interpretation of deficit is then evaluated, followed by an assessment of psychophysiological studies of schizophrenia. 4 theories of deficit are discussed: social motivation, drive, arousal, and interference. It is suggested that a fundamental sensori-motor defect underlies psychological deficit in schizophrenia. A final section describes methodological problems raised by research reviewed in both papers.

During the last 20 years a voluminous research literature has appeared on the subject of psychological deficit in schizophrenia, and the relative incapacity of this patient group has been demonstrated with a host of different laboratory tasks. In a previous paper (Buss & Lang, 1965), the authors reviewed deficit experiments concerned with concept formation, the possible disrupting character of social censure and affective stimuli, and the enhancing effect on performance of various reinforcers and motivational devices.

These findings revealed that testable schizophrenics are about as responsive to social pressure (reward or punishment) as other patients and normal controls. The hypothesis that in schizophrenia social censure invariably leads to deficit is not tenable. Regression theory interpretations of psychotic behavior receive little specific empirical corroboration.

A large body of literature indicates that affective stimuli in general disrupt the functioning of schizophrenics, but this may be due to a broader inability to inhibit any interfering stimulus. Similarly, while schizophrenics do not appear to have lost the capacity to form concepts, the concepts achieved are deviant—overinclusive or overexclusive—a flaw often traceable to the intrusion of task-irrelevant events. On the positive side, deficit can be significantly reduced by extra instructions, feedback about responses, and intense, physical reinforcers.

Guided by these considerations, the present paper is oriented around two broad conceptions: interference and activation. The first of these directly concerns schizophrenics' ability to attend to specific stimuli and to inhibit inappropriate responses. The relevant literature on attention, set, and association is explored. In general, schizophrenics show interference effects in all three of these areas. They are distracted by external stimuli. Responses are more likely to be determined by incidental physical properties of the perceptual field than by meaningful relationships. There is difficulty in initiating and maintaining a set over time and in changing a set that is no longer suitable to the experimental task. Finally, schizophrenics' associations are uncommon, intrusive, and interfere with performance.

While the first paper suggested that social motivational constructs have limited value, the fact of greatly reduced responsivity in schizophrenia argues for a thorough exploration of formal theories of motivation or activation. Experimental studies relevant to two classical drive theories are reviewed. One holds that the associational disturbance of schizophrenics is a consequence of high-anxiety drive; the other stresses reactive inhibition, which slows down learning and increases reminiscence. The drive approach is complemented by the neurophysiological concept of arousal. From this perspective, schizophrenics have been viewed both as underaroused and as so hyperactivated that effective responding is impossible. Evidence for these conceptions is evaluated in a review of research on the somatic response system in schizophrenia.

---

[1] The following abbreviations will be used: reaction time (RT), preparatory interval (PI), critical flicker frequency (CFF), galvanic skin response (GSR), electroencephalograph (EEG), electromyograph (EMG), autonomic nervous system (ANS), and stimulus generalization (SG).

[2] Now at the University of Wisconsin.

Thus, the present paper considers five specific topics: attention, set, associative interference, drive, and somatic arousal. A final discussion of deficit theory and a consideration of methodological problems follows the research review.

ATTENTION

The improvement of schizophrenics under aversive stimulus conditions has generally been explained in terms of an increase in motivation or "need" to respond. However, Lang (1959) suggested that the improvement is attributable to a stimulus-intensity dynamism and undertook a test of this hypothesis. Schizophrenics and normals performed a visual disjunctive reaction-time (RT) task under five conditions: escape, excitation, avoidance, information, and control. Following a pretest series the escape subjects were administered an aversive white noise stimulus, simultaneous with the onset of the visual cue and terminated by a correct response. Excitation subjects received the same ancillary stimulus, but the noise was terminated at random intervals unrelated to the subject's response. Escape and excitation both yielded significantly greater improvement in RT than did the control condition. Avoidance training subjects were administered aversive noise following responses slower than their pretest median. Information subjects received a nonaversive signal under the same conditions. The information group significantly lowered their RT; the avoidance procedure did not lead to significantly greater improvement than that obtained for control subjects. While the positive effects of the escape and excitation conditions did not persist into a posttest series, the positive effects of information did. These data strongly suggest that aversive stimuli improve the performance of schizophrenics via a stimulus-intensity dynamism or better definition of the task-relevant stimuli rather than via an alteration of the subject's need to respond.

Tizard and Venables (1957) found that schizophrenics improved visual RT on trials accompanied by white noise. They also found regular decreases in auditory RTs to pure tones or white noise with increases in the physical intensity of the stimulus (Venables & Tizard, 1958). King (1962b) and Grisell and Rosenbaum (1963) confirmed this result and showed that the intensity-latency gradient is steeper for chronic schizophrenics than for normal subjects. However, King reported that in both chronic schizophrenics and normals the degree of RT reduction attendant on intensity increase covaried with the subject's base latency at more moderate stimulus intensities. Furthermore, both populations showed a similar reduction in intraindividual variability with stimulus-intensity increase. King (1962b) concluded that schizophrenics and normals

> are equally sensitive to change in the variable, in a relative sense, although patient response is initially slower and thereby subject to greater reduction in absolute value [p. 304].

The effects of intensity on visual RT are not clear.[3] An initial experiment yielded variable results for different light intensities but no regular order (Venables & Tizard, 1956a). A second investigation (Venables & Tizard, 1956b) reported paradoxical effects, but these did not hold up on retest. The variable results may be due to the short intertrial period (8 seconds) used in these experiments, and possible differences between normal and schizophrenic subjects in recovery time to visual stimuli should be explored. Cohen (1949a, 1949b) reported anomalies in the eye grounds of schizophrenic patients and suggested that their visual pathways may be affected by some pathological process. In any event, withdrawn schizophrenics did yield significantly faster auditory RTs in a bright room than under conditions of dim illumination (Tizard & Venables, 1957). These authors proposed that schizophrenics suffer from a defect in the arousal system, and that

---

[3] Venables and O'Connor (1959) reported that unlike normal subjects, schizophrenics yield relatively faster RTs to light than to an auditory stimulus. Analysis of the data revealed this to be an artifact of relative speed. Slow normals also showed the light stimulus superiority, whereas fast schizophrenics had relatively better auditory RTs. However, cross modality investigations involve many procedural problems, and further study is needed before the importance of these results can be assessed.

> an increase in sensory input might be expected to have nonspecific facilitatory effects on cortical and sub-cortical activities [p. 303].

Karras (1962) attempted to test the alternate reinforcement and arousal interpretations of the deficit reduction attendant on aversive stimulation. Chronic male schizophrenics responded in a visual RT task under one of five experimental conditions. The first two groups escaped from either a low- or an intense-noise stimulus; two other groups performed while either a low- or an intense-noise was continuously administered. All four groups were compared with a no-noise control. Both escape groups had faster RTs than the two stimulation conditions, and Karras held that these findings favored a reinforcement interpretation. However, another conclusion is suggested by comparison with other data. In the previously considered experiments (Lang, 1959; Tizard & Venables, 1957) the sound was not on continuously. Tizard and Venables administered the sound only during a few trials in a long series. In the Lang experiment it was initiated at the same time as the visual stimulus and thus may have functioned both as a stimulus dynamism (arousal) and as a signal to emphasize the appropriate instant for a response (a focusing function). In the Karras (1962) experiment the continuous white noise might have aroused the subject, but it was also an irrelevant stimulus, unrelated to the task at hand. As such, it would be expected to initiate competing responses. Consistent with this interpretation is the fact that Karras's low-stimulation group yielded slower RTs than the high-stimulation group. Both the high and low stimuli were distracting, but it appears that the high-intensity sound also served a positive arousal function. Furthermore, the low-intensity stimulus was not reported as unpleasant when presented suddenly. Nevertheless, when presented coincident with the discriminative stimulus (escape), the low-intensity stimulus yielded faster RTs than the two general stimulation conditions and consistently faster RTs than the control group, although this latter difference was not significant. These effects cannot be attributed to escape from an aversive stimulus and must therefore be attributed to the increased focusing of a more marked onset change.[4]

Wienckowski (1959) also re-examined the effects of collateral stimuli on the RT task, using a modification of Lang's apparatus (1959). In addition to a control situation, the performances of acute and chronic schizophrenics and normals were evaluated under three experimental conditions: (*a*) a light and a brief buzzer initiated the preparatory interval (PI); the light alone was maintained throughout the PI and terminated with the onset of the discriminative stimulus; (*b*) both buzzer and center light continued throughout the foreperiod and were terminated with discriminative stimulus onset; (*c*) the buzzer and center light were not terminated until the subject completed his response. Under Condition *b*, both the chronic and acute schizophrenics showed a greater reduction in RT than either schizophrenic controls or normals performing under the same experimental condition. Neither Condition *a* nor *c* resulted in improved scores for any group, although the latter condition resulted in a nonsignificant lengthening of the mean RT of the chronic patients. Thus schizophrenics showed positive effects from collateral stimuli only when there was a significant net change in total stimulation at the onset of the discriminative stimulus (Condition *b*). Stimuli introduced at other times were ineffective or, in the case of the chronic group, showed a tendency to disrupt performance when they continued throughout the trial period (Condition *c*).

On the basis of their clinical observations, McGhie and Chapman (1961) attributed schizophrenic deficit to a disturbance of "selective attention," the disturbance being greatest when the patient must inhibit information in one sensory channel and attend to another. They noted that schizophrenics are particularly distracted by irrelevant auditory stimuli but that any unrelated sensory input disturbs a smooth sequence of motor responses. Substantial support for this view was obtained in an experimental study (Chapman & Mc-

---

[4] An alternate interpretation of the Lang and Karras studies is offered by Silverman (1963).

Ghie, 1962). When a sporadic, high-pitched noise was introduced, schizophrenics displayed a greater increase in errors on a visual tracking task than either normals or nonschizophrenic psychiatric patients. Similar differences were found when subjects were instructed to spin a wheel at a constant rate while listening to varying rates from a taped metronome. Shakow (1950) also reported interfering effects of distracting stimuli. Chapman and McGhie (1962) found the same effects for visual distractors: when subjects were instructed to attend only to auditory information while being simultaneously presented with competing visual cues, schizophrenics showed great disturbance of performance relative to normals or other psychiatric patients. Sutton, Hakerem, Zubin, and Portnoy (1961) studied a serial RT task and found that schizophrenics showed greater lengthening of RT to a second stimulus in a different modality than did normal subjects.

The inability of schizophrenics to exclude distracting stimuli has also been observed in cognitive tasks. Chapman (1956) had schizophrenic and normal subjects sort cards according to concepts. In addition to a "correct" figure the cards shared communalities that were not appropriate to the correct response. The schizophrenics used the incorrect distractor communalities as a basis for the sorting more than did normals. Weckowicz (1960) found that schizophrenics and brain-damaged patients performed worse than other nonschizophrenic psychiatric patients on a Hidden Figure Test, which required subjects to select relevant and disregard irrelevant information. Recently, Draguns (1963) studied a task in which subjects interpreted pictures that became progressively clearer with successive presentations. In addition to making more recognition errors, chronic schizophrenics were less able than normals to inhibit responses to the earlier, ambiguous pictures.

In addition to their inability to exclude or inhibit unwanted sensory inputs, schizophrenics have difficulty in integrating related stimuli. When this is done by the experimenter's presenting stimuli simultaneously as in the Lang and Wienckowski studies, a marked facilatory effect is observed. However, when stimuli are only slightly out of phase, the schizophrenic is unable to compensate and relative deterioration results. Chapman and McGhie (1962) found schizophrenics' performance to be markedly disturbed when they were required to repeat information presented on alternate visual and auditory channels. Schizophrenics also had greater discrepancies than other psychiatric patients between their scores on a dual task (they were instructed to indicate which figures appeared in a series twice *and* locate their position) and performance when instructed to do only one of these tasks.

Distractibility and disrupted information processing may underly schizophrenic deficit in a variety of perceptual tasks. Cohen, Senf, and Huston (1956) and Johannsen, Friedman, and Liccione (1963b) studied pattern recognition in schizophrenics, using the Street Visual Gestalt Test. They found inferior closure in chronic patients relative to both acute schizophrenics and normal controls. Eysenck, Granger, and Brengelmann (1957) reported a lower than normal (although not significant) mean Street Test score for a small group of psychotics; a significant closure deficit was found on a Kohs blocks task when these patients were compared to either normals or neurotics. Snyder, Rosenthal, and Taylor (1961) reported a closure deficit in schizophrenics, based on drawings of incomplete figures. However, Snyder (1961) found a greater than normal tendency to closure in a group of acute, paranoid patients, which he attributed to emotionality associated with disease onset. Johannsen et al. (1963b) reported a significant interaction between the paranoid-nonparanoid dimension and chronicity. Their paranoid subjects maintained normal closure test scores except for the most extreme chronic group, but nonparanoid schizophrenics showed progressive deterioration with chronicity.

Although evidence is limited, a few studies suggest differences between schizophrenics and normal controls in critical flicker frequency (CFF). Dillon (1959) reported higher CFF thresholds in functional psychotics than in a control group, but the patients were a specially selected treatment sample. McDonough (1960) also found higher than normal mean CFFs in reactive schizophrenics but no differences between process schizophrenics and normals. Recently, Johannsen et al. (1963b) reported that acutes showed a high CFF, but with increasing chronicity there was a progressive reduction in thresholds to a point well below that of normal subjects. The conflicting findings may be due to the competing effects of high distractibility and high activation level in schizophrenia. These two factors tend to influence CFF thresholds in opposite ways. The most extensive study (Johannsen et al., 1963b) is consistent with the hypothesis of decreased vigilance in chronic patients.

A number of researchers report deficit in size or distance constancy in chronic schizophrenics (Cooper, 1960; Crooks, 1957; Lovinger, 1956; Reynolds, 1954; Weckowicz, 1957, 1958, 1960; Weckowicz & Blewitt, 1959; Weckowicz & Hall, 1960; Weckowicz, Sommer, & Hall, 1958). Nevertheless, Reisman (1961) failed to find differences in size constancy between normals and two schizophrenic groups. Raush (1952, 1956) found that "fairly young, fairly well oriented" paranoid schizophrenics showed higher constancy scores than normal subjects. Sanders and Pacht (1952) also reported what amounts to "overconstancy" (distant objects seen as larger than they actually were) in a group of psychotic outpatients. However, Kidd (1964) found subnormal monocular depth perception in schizophrenics within 48 hours after they were admitted to a state hospital. Unfortunately, it is not clear whether these were first admissions, nor is there information on the subtype or chronicity of these patients. Perhaps the most comprehensive data on these variables is provided by Johannsen et al. (1963b) who studied depth perception in a large group of schizophrenics classified according to duration of illness. They found a low mean error score for acute patients; chronic patients generally made significantly more errors than either acute psychotics or normal controls, although the most extreme chronic group evidenced some restitution of ability.

Thus, the closure, size constancy, and distance constancy studies reviewed here found deficit in schizophrenics more often than not. However, experiments with acute schizophrenics, and particularly with early paranoid patients, suggest the presence of compensatory efforts in these subjects, which result in normal or occasionally above-average functioning. Chronic patients show the greatest deterioration, but even this result is not wholly reliable.

Perhaps much of this variability between experiments reflects the schizophrenic's response to characteristics of the specific task or task setting that do not affect normal subjects. The perceptual judgments of schizophrenics appear to be readily manipulated by minor alterations of experimental stimuli. Weckowicz (1964) found inferior shape constancy in schizophrenics when the stimulus objects were inclined at 60 degrees, but not when the angle of inclination was only 30 degrees. The findings of Raush (1956) and Pishkin, Smith, and Leibowitz (1962) among others, show significant constancy differences attributable to the symbolic value of stimulus objects. It has already been suggested (Buss & Lang, 1965) that these results are due to the greater number of associations occasioned in schizophrenics by these stimuli and their intrusion into the task at hand. Salzinger (1957) reported no initial differences between schizophrenics and normals in ability to discriminate weights. However, when required to make estimates following the lifting of a heavy anchor stimulus, the patient group showed a significantly greater shift in judgment than controls. Lovinger (1956) found deterioration in the size constancy of schizophrenics only under minimal cue conditions. Leibowitz and Pishkin (1961) reported no size constancy deficit in schizophrenics under maximum cue conditions, but a follow-up investigation (Pishkin et al., 1962) revealed a significant difference in size estimation errors between normals and chronic schizophrenics. In this latter experiment the subjects had to illuminate the stimulus objects themselves by closing a switch; in addition to making more errors, schizophrenics tended to illuminate the field significantly more often than controls. Pishkin and his associates attributed these differences to the schizophrenics' greater difficulty in selecting and attending to relevant cues. Further evidence for an attention deficit is provided by tachistoscopic studies, e.g., McGinnies and Adornetto (1952) found that schizophrenics have higher recognition thresholds than normal for *both* emotionally toned and neutral verbal material.

Schizophrenics appear to live in a subjective world determined by the physical size and intensity of stimuli, poorly integrated with the information about relationships between objects, which modulates the perception of normal subjects. Perhaps because attention is so wavering, the static, formal characteristics of the physical field determine behavior, and schizophrenics fail to show normals' flexible, meaningful perceptual experience. Thus, deficit is greatest when the schizophrenic must pay attention to more than one stimulus input, switch his attention from one stimulus to another, or ignore irrelevant stimuli in favor of physically weak, task-relevant inputs. Deficit is least when the irrelevant stimuli are few, the task stimulus is intense and unequivocal, or any collateral inputs are temporally in phase. Under these latter conditions chronic schizophrenics may approach the functioning level of normal subjects.

SET

Analogous to the difficulty with concurrent stimuli described above is the schizophrenic's frequent failure to inhibit inappropriate responses initiated by sequences in time. This was already suggested by the Chapman and McGhie (1962) experiment in which patients failed to spin a wheel at a constant speed while listening to varied tempi from a metronome. A number of studies indicate a time estimation deficit in schizophrenia. Rabin (1957) found schizophrenics to be significantly poorer than nonpsychotics in estimating the length of a psychological interview. Lhamon and Goldstone (1956), H. E. King (1962a), and Pearl and Berg (1963) all reported that schizophrenics systematically overestimate short time intervals (.5 to 30 seconds). King's interval-matching study indicated that normal subjects consistently underestimate the same intervals, and he suggested that the deficit of schizophrenics may be attributed to their characteristic psychomotor retardation. Pearl and Berg (1963) had schizophrenics and normals estimate the presentation time of neutral and affect arousing pictures and found that psychotics displayed greater overestimation to the affective materials.

The schizophrenic's performance is markedly disturbed whenever he must initiate a response at a fixed point in time. Shakow (1962) has noted that schizophrenics show less deficit on tasks such as tapping, in which time of onset and termination is not important to adequate performance, than on tasks like RT, which require preparation for a response at the discretion of the experimenter. Furthermore Rosenthal, Lawlor, Zahn, and Shakow (1960), and H. E. King (1954, 1961) have demonstrated a high positive relationship between RT and mental health ratings.

The latencies of schizophrenics are readily manipulated by slight variations in the foreperiod or preparatory interval (PI). The RT of chronic patients increases disproportionately with the lengthening of a regular PI (Huston, Shakow, & Riggs, 1937; Rodnick & Shakow, 1940). Zahn, Shakow, and Rosenthal (1961) demonstrated that this effect is attributable not to the slower pace of the task but to foreperiod length; it is also partially a function of the preceding sequence of foreperiods. When a single long PI occurs in a series of shorter ones, the RT latency is relatively shorter than if the preceding foreperiods were of equal length (Rodnick & Shakow, 1940; Tizard & Venables, 1956). Simiarly,

> when a single long PI trial was followed by a single short PI trial, RT on the latter trial was disproportionately lengthened in schizophrenic as compared with normal subjects [Zahn, Rosenthal, & Shakow, 1961, p. 167].

Zahn et al. (1961, 1963) demonstrated that if the length of the PI is systematically varied during the task, the RT of schizophrenics changes in characteristic ways. When successive PIs were administered in order of increasing length, RT was an increasing function of foreperiod length for both schizophrenics and normals, but the slope of the curve was significantly greater for the former population. The RTs of the schizophrenics tended to be the same for all PIs and relatively slow when a descending order of administration was employed, whereas the curve for normal subjects was virtually identical to that when length was increased.

In summary, the RT of schizophrenics is controlled to a great extent by the context of PIs and more particularly by the immediately preceding foreperiod. As with concurrent stimuli, temporal sequence may be manipulated to produce increases or decreases in deficit.

Blaufarb (1962) demonstrated a facilatory effect of context on a verbal task. Schizophrenics and normals were administered proverb tasks in two sessions. In one session they were asked to give the meaning of a single proverb, and in the other session they were asked to give the meaning of a set of proverbs, all of which had the same meaning. The chronic schizophrenics were significantly poorer than normals under the single proverb condition. Schizophrenics showed significantly greater improvement with proverb sets and under this condition could not be distinguished from the normal group.

However, tasks may be selected so as to produce inappropriate sets, thereby increasing deficit. Mandl (1954) found that paranoid schizophrenics showed greater rigidity than normals on a perceptual task. Shakow (1950) noted that distractors disrupted the performance of both normals and schizophrenics, but the deleterious effects carried over to subsequent nonstress trials only for the psychotic patients. In W. O. Smith's (1959) study of the pursuit rotor task, process schizophrenics showed less improvement with modification of the experimental conditions than did reactive patients. Similarly, Crumpton (1963) showed that schizophrenics persisted in a previously rewarded but now incorrect response for significantly more trials than a normal control sample. Furthermore, this tendency to persist with a maladaptive response appeared to be related to the severity of the schizophrenic process.

ASSOCIATIVE INTERFERENCE

It was Bleuler (1950) who first pointed to association as the crucial issue in schizophrenia. He believed that many schizophrenic symptoms (hallucinations, delusions, etc.) were only elaborations of, and secondary to, the primary disturbance in association. This disturbance in association may be seen in bizarre ideas, loose associations, fragmented thinking, and the blocking of the usual and common chains of associations and ideas.

What is the role of associative disturbance in schizophrenic deficit? Specifically, how would difficulty in association disturb attention or set and therefore worsen performance on psychomotor, perceptual, and learning tasks? The answer may be framed in the context of the psychological demands of such tasks. The subject is required to maintain a state of vigilance, a readiness to respond to oncoming stimuli. He must also react to stimuli and instructions in a normative manner, that is, on the basis of the common meanings of stimuli and instructions. To the extent that stimuli have individual, idiosyncratic meanings (associations) for the subject, he will be handicapped in responding to them. For example, stimuli seen as relatively neutral by normals, may be provocative for schizophrenics, who perceive simple stimuli in idiosyncratic ways (Feldman & Drasgow, 1951).

In maintaining a state of vigilance as preparation for oncoming stimuli, the subject needs to be free from distractions. He must inhibit responses not only to inappropriate external stimuli but also to extraneous thoughts and associations that may divert his attention. The research on psychomotor performance discussed earlier shows that the schizophrenic is unable to accomplish the first of these tasks. Clinical reports suggest that he also cannot prevent the intrusion of bizarre ideas and associations that pull his attention from the task at hand. The idiosyncratic associations of schizophrenics distract them, thus degrading their performance. This formulation leads to two testable hypotheses.

The Associations of Schizophrenics Are Uncommon

Several studies have sustained this assertion. Moran (1953) tried a word association task with schizophrenics and normals, the stimulus words having previously been defined by the subjects. The schizophrenics' associations were significantly less related to the stimulus words than the associations of the normals. These findings were corroborated by Johnson, Weiss, and Zelhart (1964) who found that schizophrenics produced more idiosyncratic word associations than normals.

Sommer, Dewar, and Osmond (1960) compared the associations of schizophrenics and normals to the words on the Kent-Rosanoff list. Schizophrenics gave significantly more uncommon associations, and they were extremely variable in terms of both stability of responses over time and comparisons between subjects. Sommer, Witney, and Osmond (1962) followed up this experiment by trying to condition common associations. While alcoholics conditioned rapidly, schizophrenics showed very little conditioning.

Wynne (1963) found that the free associations of acute schizophrenics differed little from those of normals. However, when instructed to give the associations "most people do," normals gave more common associations but schizophrenics did not. Maltzman, Seymore, and Licht (1962) attempted to condition normal subjects to give common or uncommon associations. They learned to give more common associations but did not learn to give more uncommon associations. Maltzman, Cohen, and Belloni (1963) found that schizophrenic children give more uncommon associations than normal children.

On the basis of these various studies, we may conclude that schizophrenics tend to give uncommon associations but cannot learn common associations, whereas normals tend to give common associations but cannot learn uncommon associations.

Intrusive Associations Worsen the Performance of Schizophrenics More Than of Normals

Chapman (1958) used a verbal concept formation task, with a stimulus word and three response words. One response was the correct concept, one an (irrelevant-to-the-task) association, and one neither. The associations were of high and low strength. Schizophrenics showed significantly more associative intrusions than normals.

Donahoe, Curtin, and Lipton (1961) questioned whether associative intrusions in schizophrenics occurred only with meaningful stimuli (words) or with all stimuli (words and nonsense syllables). They tested the effect of experimentally built-in intrusive associations on the learning of nonsense syllables and found an equal decrement in normals and schizophrenics. Downing, Ebert, and Shubrooks (1963) studied the number of errors schizophrenics produced on a concept test by different types of distractor words. Associatively linked words produced more errors than either contiguity or rhyme-clang distractors. Thus it is the restrictive hypothesis that appears to be correct: schizophrenics suffer especially from the intrusion of meaningful irrelevant associations and not all types of irrelevant associations.

A number of studies have produced more direct evidence on the intrusion of associations in schizophrenia. Lang and Luoto (1962) had subjects learn two lists of paired associates (words). On the second list, half the response terms were associates of the response terms used in the first list. The response terms of the other half were also associates, but they were not assigned to the correct stimulus term, thereby creating an interference list. The results showed that the schizophrenics' mediational processes were not significantly different from the normals. However, schizophrenics showed significantly poorer performance than normals on the early trials of the interference list. Furthermore, while normals seldom again offered the response terms of pairs already learned, these words tended to persist as responses for schizophrenics, interfering with subsequent learning.

Spence and Lair (1964) failed to find differences between the paired-associate learning of normals and schizophrenics, consistent with Mednick's theory (1958). However, an analysis of errors did differentiate the two groups. Acute schizophrenics tended to give more overt, inappropriate responses, stemming from both intralist and extralist sources. Normals produced more errors of omission. Thus, schizophrenics seem unable to inhibit the overt expression of intruding associations.

Further evidence of intrusiveness comes from an experiment by Lauro (1962), who used words that varied in clustering tendency. There were no differences in recall between normals and schizophrenics in the easy clustering list, but in the difficult clustering list the schizophrenics "imported" more irrelevant words than did normals.

There should be more associative interference when words have several meanings than when they have only one meaning. Faibish (1961) had normals and schizophrenics define and free associate to words with either one meaning or multiple meanings. Both normals and schizophrenics showed poorer word association and vocabulary performance with the multiple-meaning words, but the schizophrenics' decrement was greater than the normals. The schizophrenics were disrupted by the multiple meanings, and Faibish concluded that "the majority of the results can be understood in terms of interference effects [p. 423]."

Finally, Lester (1960) studied restricted association in normals, hebephrenic and paranoid schizophrenics, and epileptics. The patient groups showed more interference than the normals in the selection of associates, the interference occurring because of the intrusion of extraneous associations. The interference was greatest for epileptics, followed by hebephrenics, paranoids, and normals, in decreasing order.

In brief, the hypothesis of associative interference has been verified. Schizophrenics have more unique, nonshared associations, and these associations, like external distractors, serve to deteriorate performance because of their intrusive nature. Shakow's (1962) summation of his RT studies aptly describes both inner (associational) and outer distractors:

> Here we see particularly the various difficulties created by *context*, the degree to which the schizophrenic is affected by irrelevant aspects of the stimulus surroundings—inner and outer—which prevent his focusing on the "to-be-responded-to" stimulus. It is as if, in the scanning process which takes place before the response to a stimulus is made, the schizophrenic is unable to select out the material relevant for optimal response. He apparently cannot free himself from the irrelevant among the numerous possibilities available for choice. In other words, that function which is of equal importance as the response *to* stimuli, namely, the protection *against* the response to stimuli, is abeyant [p. 25].

DRIVE

Two drive theories of schizophrenic deficit have been proposed, each associated with different behaviors. Positive drive theory, which identifies anxiety as the crucial variable, is intended to explain aspects of the schizophrenic's associational disturbance. Negative drive theory focuses on inhibition during conditioning and the phenomenon of reminiscence.

Anxiety and Associational Deficit

The hypothesis of continuity between neurosis and psychosis has long been popular with theorists and practicing clinicians. From this perspective, the symptoms of schizophrenia result from a failure of neurotic defenses. The individual is finally overwhelmed by social or personal anxiety, and he retreats to the pseudoworld of psychosis. Fenichel (1945) and Arieti (1955) provide classical examples of this approach to clinical phenomena.

Mednick (1958) has recently attempted an analogous, though more rigorous, explanation of the schizophrenic's associational disturbance. Anxiety is again taken to be the central construct, but it is equated with the Hull-Spence concept of drive.

The term *drive* is frequently used to denote physiological activation; it is not considered here in that context (see the following section on arousal). In the Hull-Spence framework, drive is generally defined by deprivation time or stimulus intensity and measured by performance. As an intervening variable it bears a mathematical relationship to response strength, learning, and generalization; Mednick's hypotheses are deduced within this theoretical system. Experimental evidence is sought at a number of crucial points, and the following review is concerned mainly with this empirical support.

Mednick (1958) assumes anxiety to be intense in the schizophrenic. High drive (intense anxiety) leads to excessive stimulus generalization and associative generalization. A phobic defense may be successful in maintaining a precarious balance, as seen in schizoid, withdrawing personalities, but the balance may be upset by a precipitating event, which elevates the anxiety level. This increase in drive leads to still more generalization, and now many more stimuli are fear-invoking. Furthermore, the higher drive level increases the intensity of previously present fear responses.

> As the spiral of anxiety and generalization mounts, his drive level may increase to an almost insupportable degree. As this is taking place, his ability to discriminate is almost totally eclipsed by his generalization tendencies. Any unit of a thought sequence might call up [still another] remote associate. . . . Clang associates based on stimulus-response generalization may be frequent. . . . His speech may resemble a "word salad." He will be an acute schizophrenic with a full-blown thinking disorder [p. 322].

Mednick recognized that, in contrast to acute schizophrenics, chronic schizophrenics tend not to give overt evidence of intense anxiety. Therefore he proposed the following transition from acute to chronic schizophrenia. The excessive generalization of the high drive (anxiety) state may lead to a "highly generalized, remote, irrelevant, tangential associate." A remote association diverts the individual's attention from anxiety-provoking stimuli, and the resulting drive (anxiety) reduction is reinforcing. Continued repetition of the strongly reinforced tendency to escape anxiety via remote associations leads to deviant, disorganized thinking. Thinking irrelevant thoughts proves to be so effective in reducing anxiety that the schizophrenic may appear emotionally phlegmatic. Now the well-learned tendency toward remote and tangential associations is maintained even in the absence of a high anxiety level. In deriving his theory Mednick made four assumptions.

Schizophrenics acquire classically conditioned responses faster than normals. Two studies support this assumption. Pfaffman and Schlosberg (1936) demonstrated more frequent conditioned patellar tendon reflexes in schizophrenics than in normals. Spence and Taylor (1953) reported similar results for eyelid conditioning.

On the other hand, six studies have shown that normals condition at least as well as, or better than, schizophrenics. Shipley (1934) and Pishkin and Hershiser (1963) found better conditioning for normals with the GSR, and Howe (1958), also using GSR conditioning, failed to obtain significant differences between normals and schizophrenics. Franks (1954), Peters and Murphree (1954), and O'Connor and Rawnsley (1959) also found that schizophrenics fail to learn faster or better than normals on a variety of conditioning tasks. Thus the weight of evidence is against the hypothesis, which receives little support from the studies just cited.

In more complex situations schizophrenics learn slower. This assumption is linked to the first one in an attempt to apply drive theory to two types of learning situations. In simple situations (e.g., classical conditioning) high drive leads to faster learning; in complex situations high drive enhances irrelevant and incorrect responses, causing a decrement in learning. Schizophrenics, having greater anxiety and therefore higher drive, should learn faster in simple situations and slower in complex situations. The second part of this statement is true: schizophrenics do learn slower in complex situations. However, the first part is probably untrue. Except for classical conditioning, where the evidence is admittedly equivocal, schizophrenics learn *slower* than normals in simple situations. (See the sections on Insufficient Motivation, Attention, Set, and Associative Interference.)

Actually the facts already reviewed are a good deal more complex than the hypothesis allows. While chronic schizophrenics rarely show superior learning to normals, they may approach normal functioning when the response alternatives are reduced and distractions are few. If such conditions define task simplicity, these facts may bring the drive theorist some solace. Furthermore, in a few experiments acute paranoid patients have actually emerged superior to their normal controls. Unfortunately, one must pick and choose among the data to find support for what pretends to be a comprehensive theory. Since schizophrenics in general learn both simple and complex tasks more slowly than normals, the evidence is not consistent with drive theory.

Schizophrenics overgeneralize in comparison to normals. Mednick cited four experiments that ostensibly found elevated generalization gradients in schizophrenics. The first is a study by Bender and Schilder (1930), who present data that are difficult to interpret in terms of generalization gradients and who did not employ normal controls.

The second is Garmezy's study (1952), which is open to several interpretations, as we showed in the previous paper (Buss & Lang, 1965). The third is Mednick's (1955) doctoral dissertation, in which he wrote:

> With respect to the hypothesis that schizophrenics would display a more elevated GSR (gradient of stimulus generalization) than normals, the results are not conclusive. While the C (normal) and S (schizophrenic) groups differ, the differences occur both in the predicted direction and counter to it [p. 540].

The fourth study is that of Dunn (1954) who found an elevated generalization gradient in schizophrenics in one experimental condition and no differences between the gradients of schizophrenics and normals in three other conditions. The most neutral comment to be made about these four studies is that they do not offer support for this third hypothesis.

There is considerable evidence that schizophrenics tend to be overinclusive, including in their vocabulary definitions and conceptual sortings more stimuli than do normals (Buss & Lang, 1965).

> Phenomena of this sort tempt one to describe schizophrenics as showing a heightened and broadened gradient of secondary stimulus generalization. However, the experimental evidence seems to contradict this interpretation. At least three different investigators have compared schizophrenics and normals on tasks which are usually thought of as measuring propensities toward heightened semantic generalization. All of these studies have uniformly obtained negative results, that is, they find no difference between schizophrenics and normals on semantic generalization [Chapman, 1962].

Chapman went on to report an experiment demonstrating that schizophrenics were *less* inclusive than normals. Thus the assumption that schizophrenics overgeneralize is not supported by the evidence.

High anxiety leads to overgeneralization and to faster conditioning. This assumption receives some support from the literature, which is too remote from schizophrenic deficit to review here. It should be noted that research with the Taylor Manifest Anxiety scale (the major instrument in these studies) has tended to produce controversial and, at times, unreliable results. Mednick himself (1957) found that middle-anxious subjects had flatter stimulus generalization (SG) gradients than did high- and low-anxious psychiatric patients. Thus there has been both positive and negative evidence on the relationship between anxiety and overgeneralization.

The results concerning anxiety and classical conditioning have been more clearly positive: high-anxious subjects do condition faster than low-anxious subjects. However, the potency of anxiety as a determiner of conditioning is slight, a fact that is admitted by drive theorists:

> While previous studies have demonstrated a relationship between conditioning and manifest anxiety, variously defined, correlation coefficients that have been reported between those two variables indicate that a relatively small amount of the variability among *S*s can be accounted for in terms of anxiety [Taylor & Spence, 1954, p. 502].

Thus the evidence of this fourth assumption tends to be equivocal.

Examination of Mednick's four assumptions indicates that supporting evidence is either equivocal or lacking. Since these assumptions are the base of his drive theory, the theory itself is weakened to the extent that the assumptions lack verification.

The theory places all its eggs in one basket in that it accounts for schizophrenia solely in terms of anxiety. The difficulty is that, with anxiety so prevalent, it is necessary to explain why schizophrenia is still relatively rare in the population. Mednick (1958) was aware of this issue of "over-explanation:"

> Why doesn't everybody proceed to schizophrenia after an extremely anxiety provoking event? The answer lies in three factors: the individual's original drive level, his rate of recovery from anxiety states, and the number of stimuli that elicit anxiety responses from the individual . . . high drive, slow recovery rate, and the number of fear arousing stimuli are highly correlated factors [p. 323].

Only those with all three factors tend to become schizophrenic, but Mednick himself admitted that the three factors are all highly correlated. Thus an extremely anxious individual, being high on all three factors, should become schizophrenic. Clearly, Mednick has not answered his own question. We may guess that he cannot answer it because he identifies anxiety as the sole cause of the thinking disorder in schizophrenia. This explanation will not be acceptable to the majority of psychologists. It is evident that many individuals with extremely high levels of anxiety never become schizophrenic, whereas Mednick's drive theory clearly implies that they should.

Reactive Inhibition and Reminiscence

Pavlov (1941) suggested that in schizophrenia inhibition predominates over excitation, the theory being labeled "protective inhibition." Several English investigators have developed this idea in terms of a negative drive, reactive inhibition (Claridge, 1960; Eysenck, 1961; Rachman, 1963; Venables & Tizard, 1956c). Thus schizophrenics are believed to develop reactive inhibition faster than normals and to dissipate it slower. If this is true, schizophrenics should learn more slowly than normals, and, more important, show greater reminiscence. The prediction is straightforward: after a rest period schizophrenics should have a greater increment in performance (reminiscence) than do normals.

This prediction has been unequivocally confirmed in only one study. Huston and Shakow (1948, 1949) tested schizophrenics and normals on a pursuit rotor task at 3-month intervals. Normals performed significantly better than schizophrenics, but whereas normals showed no reminiscence after a 3-month interval, schizophrenics manifested a clear improvement in performance.

Venables and Tizard (1956c) used a repetitive choice task and a 1-minute rest period. Schizophrenics showed slightly more reminiscence than psychotic depressives; there were no normal subjects. Bleke (1955) used a memory task under conditions of reward or punishment. With verbal punishment, poor premorbid schizophrenics showed more reminiscence than either good premorbid schizophrenics or normals. With verbal reward, there were only chance differences in reminiscence. Smock and Vancini (1962) obtained *less* reminiscence in schizophrenics after censure; after reward there were no differences in reminiscence between schizophrenics and normals. Higgins and Mednick (1963) had early and advanced schizophrenics copy the alphabet upside down and backward. They found more reminiscence in early than in advanced schizophrenics. The discrepancies among these last three studies may be due to the way punishment was administered, the difference in tasks, or sampling differences in the subjects.

Taken as a whole, these studies offer little support for the hypothesis that schizophrenics build up reactive inhibition faster than normals and therefore show greater reminiscence. Six experiments have yielded clearly negative results. Campbell (1957) used paper-and-pencil mazes, Rosenbaum, Cohen, Luby, Gottlieb, and Yelen (1959) employed the pursuit rotor, and Venables (1959) used a repetitive choice task; none of these workers found any difference in reminiscence between normals and schizophrenics. O'Connor (1957), Claridge (1960), and Rachman (1963), with varying tasks and rest periods, all found greater reminiscence in normals than in schizophrenics.[5]

In summary, there appear to be no consistent reminiscence differences between schizophrenics and normals. Some of the discrepancies among results are undoubtedly due to variations in task, in rest period, and in composition of the subject samples. Regardless of methodological considerations, it seems safe to conclude that the negative drive hypothesis has not been corroborated.

SOMATIC AROUSAL

Autonomic, cortical, and neuromuscular response systems are of considerable relevance to both activation and interference interpretations of schizophrenic deficit. The basic concept of an arousal system in the lower brain was initially dependent on electroencephalogram (EEG) findings rather than on overt behavioral data. Malmo (1958) has stressed the importance of peripheral physiological responses (GSR, EMG, cardiac rate, and blood pressure) as concomitant estimates of drive or the aroused state. Furthermore, the work of Lacey and Lacey (1958a), Gellhorn (1957), and Barratt (1962) indicates that feedback from autonomic nervous system (ANS) activity and muscle tension may have important effects on cortical functioning and thus determine temperament, alertness, and adequacy of psychomotor control. It has been suggested that schizophrenic deficit is attributable to *diminished* feedback from peripheral sensors or effector systems, with a resultant disturbance of control and orientation. Others propose that *excessive* feedback of ANS activity interferes with organized behavior in a manner analogous to that of irrelevant associations or intrusive external stimuli.

---

[5] Rachman studied length of the intervening time period and amount of reminiscence. He found that schizophrenics show greater reminiscence after 24 hours than they do after 10 minutes. Furthermore, a comparison with data published elsewhere (Rachman, 1962) suggests that the degree of reminiscence, when measured at the first trial following a 10 minute rest period, is less for schizophrenics than for normal subjects. The relationship between time interval and reminiscence is complex and in general outside the scope of the present review. The reader is referred to the Rachman papers for a discussion of these issues and the related topic of disinhibition.

Research on these covert responses in schizophrenics may be divided into two broad areas: *habitual levels of activity* and *reactivity*. The first term refers to the base amount of ANS or neuromuscular responding characteristic of an individual at rest, when external stimuli are abeyant. The second term refers to the form or amplitude of responses to specific stimuli introduced by the experimenter.

Habitual Level of Activity

At the time Hunt and Cofer (1944) reviewed literature on psychological deficit, no consistent differences between normal and schizophrenic subjects in resting ANS response level had been found (Freeman & Pathman, 1943). A later review (Hoskins, 1946) noted that schizophrenics were less variable than normals in blood pressure and the oral-rectal temperature differential.

Early investigators sought unsuccessfully for evidence of ANS hypoactivity to parallel the clinically observed withdrawal and flattened affect. In addition to this theoretical bias they were hampered by a tendency to consider single measures of specific physiological functions—heart rate or skin resistance—both as adequate estimates of overall ANS functioning and as indicants of general emotionality or arousal. However, the experimental literature (Lacey & Lacey, 1958b) suggests that the relationships between somatic responses are complex and that more than one system must be considered in such evaluations. In the subsequent few paragraphs studies of basal skin resistance, cardiovascular and respiration responses, and muscle activity in schizophrenia are reviewed.

Basal skin resistance. Hock, Kubis, and Rouke (1944) note that withdrawn psychotics show increasing skin resistance over time periods in which normals are relatively stable. Experiments by Jurko, Jost, and Hill (1952) and Howe (1958) found higher resting resistance levels for schizophrenic patients than for normal controls. Malmo and Shagass (1949), DeVault (1957), Ray (1963) in a study of female patients, and Pishkin and Hershiser (1963) reported no difference in resistance between normals and schizophrenics. Only two studies found low basal skin resistance in schizophrenics. Zahn, Rosenthal, and Lawlor (1963) report both lower than normal skin resistance and greater spontaneous activity in schizophrenics. Williams (1953) also found lower than normal resting levels for what he described as an "early chronic schizophrenic" group. However, this difference barely reached significance and was not maintained during the experimental conditions. Previous studies by Syz (1926) and Syz and Kinder (1928) attempted to distinguish between patient subtypes. They found high basal resistance for catatonics, the paranoids' mean fell between those obtained from two normal control samples, and there was more spontaneous activity in paranoids than catatonic patients.

Cardiovascular system and respiration. Gunderson (1953), Williams (1953), and Jurko et al. (1952) reported faster resting heart rates for schizophrenics than normal controls. DeVault (1957) obtained similar results for chronic, reactive patients but not for process schizophrenics. However, Reynolds (1962) found process schizophrenics had a significantly higher basal pulse than reactive patients, and normal controls yielded the lowest rates.

Altschule and Sulzbach (1949) reported vasoconstriction, particularly of the hand, to be a habitual condition in many schizophrenics. Data collected by Henschel, Brozek, and Keys (1951) also suggested that the skin vessels of schizophrenics have a high resting tonus level. Consistent with these views, Malmo and Shagass (1952) found habitual high diastolic pressure in chronic schizophrenics, combined with lowered pulse pressure. Both diastolic and systolic pressures tended to be higher in process schizophrenics than normal subjects in the Reynolds (1962) experiment, but a significant difference was obtained for only one of four resting samples.

Gunderson (1953), Jurko et al. (1952), and Williams (1953) reported that schizophrenics had faster respiration rates than normals. Reynolds' data show a nonsignificant tendency for process schizophrenics to have faster rates.

Muscle activity. Malmo and his associates (Malmo, & Shagass, 1949; Malmo, 1950; Malmo, Shagass, & Smith, 1951) reported a high resting electromyograph (EMG) in schizophrenics, with the highest levels among chronic patients. Martin (1956), Whatmore and Ellis (1958), and Petursson (1962) also found evidence of higher than normal muscle tension in schizophrenics. Reynolds (1962) reported a significantly higher resting EMG response in process schizophrenics than normal subjects; reactive patients fell between these two groups. Jurko et al. (1952) found considerable adventitious muscle activity in schizophrenics relative to normals. Malmo and Shagass (1949), Edwards and Harris (1953) and later Gindis (1960) reported disturbance of finger movement in a variety of schizophrenics.

EMG investigations parallel psychomotor studies (King, 1962b) in that schizophrenics show high intraindividual variability. Malmo, et al. (1951) found that increases in painful stimuli yielded increased action potentials in the neck muscles of schizophrenic patients, but unlike normals, these subjects failed to show such changes in potentials taken from arm electrodes. Reynolds (1962) also reported high variability in response to stressors and within and between subjects during rest.

The psychophysiological studies described above generally included only one testing session, and little information is available on trends in variability. Recently Carrigan (1963) reported that daily polygraph tests yielded few differences in intraindividual variability between a group of nonparanoid schizophrenics and normal controls. Acker (1963) compared a small group of schizophrenics on tranquilizers and normal subjects. The two groups yielded different trends for different physiological systems. Schizophrenics showed heart rate adaptation over sessions, but unlike the normal controls, blood pressure measures did not show this effect. Both psychophysiological variability and adaptation effects in schizophrenia deserve further study.

In summary, the habitual level of somatic activity in schizophrenics appears to have the following characteristics. Skin resistance levels are generally similar to those of normals, although two samples were clearly higher and at least one was lower than those of control subjects. During experiments the cardiovascular systems of schizophrenics tend to be at a higher activation level than those of normal controls. However, the relationship between cardiac functioning and such dimensions as process-reactive or chronic-acute is not yet clear. All reports indicate higher than normal muscle tension in schizophrenics, the highest levels being associated with chronicity and the process label.

The conclusion that schizophrenics are underaroused or at a normal level of arousal during experiments (i.e., Ray, 1963) is not consistent with a major part of the data. Only the skin resistance findings lend support to this position, while the cardiovascular and neuromuscular systems point to a heightened level of activation. The reason for this division has been variously interpreted. Jurko et al. (1952) point out the close relationship between sweat gland activity, attention, and ideation—functions particularly disturbed in schizophrenics. They suggest that energy discharge via muscular tension represents a phylogenetically more primitive way of maintaining energy balance than through the electrodermal response.

The skin resistance findings seem inconsistent with Wenger's (Wenger, 1956; Wenger, Jones, & Jones, 1959) conclusion that the autonomic activity of schizophrenics is dominated by the sympathetic system. However, Solomon, Darrow, and Blaurock (1939) remind us that sweat gland activity, though sympathetically activated, is a cholinergic mechanism. They suggest that the neurohormonal inhibition of cholinergic response systems may be related to psychotic withdrawal. More recently, Rubin (1962) has suggested that an adrenergic-cholinergic unbalance is an important aspect of functional psychosis. Perhaps sweat gland activity is a sensitive measure of small increases in arousal level, but the action of adrenin inhibits this system under high levels of sympathetic activation.

The cardiovascular results and particularly the muscle tension findings suggest that the level of "biological noise" is quite high in schizophrenics. A number of authors have proposed that this directly accounts for schizophrenic symptoms. Angyal (1935, 1936) suggested that disturbance of muscle tension may be the perceptual basis of somatic delusions in schizophrenia, and Gould (1950) proposed that auditory hallucinations could be traced to a motor disturbance of the speech mechanism. Furthermore, diffuse neuromuscular activity may contribute to deficit in psychomotor or perceptual tasks by interfering directly with coordinated behavior (Freeman, 1948; Luria, 1932). Wishner (1955, 1962) has maintained that an increase in degree of psychopathology is signaled by a decrease in efficiency, with efficiency defined as the ratio of focused to diffuse activity. Normals orient their activity to the task requirements, and there is a minimum of diffuse random activity in either the musculature or the ANS. In psychopathology less of the total activity is directed to the task at hand, and more of the total behavior output is diffuse and random. Schizophrenia represents the extreme of inefficiency.

Reynolds (1962) found that in process schizophrenics stress sometimes produces a decrease rather than the normal's increase in muscle tension level. He interpreted studies which show chronic schizophrenics improve under conditions of aversive stimulation (Cohen et al., 1956; Lang, 1959) in a manner that parallels Wishner's theory. Reynolds suggested that aversive stimuli reduce the general tension level in chronic schizophrenics, producing a temporarily favorable ratio of directed to diffuse muscular activity and an associated amelioration of psychomotor performance.

Reactivity and pathology. Many studies of reactivity to stimulation reveal a diminished ANS response in chronic schizophrenia. Stressors that have yielded hyporeactivity include the inhalation of heated air (Freeman & Rodnick, 1940) and cold baths (Buck, Carscallen, & Hobbs, 1950). Schizophrenics show reduced rotational and caloric nystagmus (Angyal & Blackman, 1940, 1941; Angyal & Sherman, 1942; Colbert, Koegler, & Markham, 1959; Freeman & Rodnick, 1942; Leach, 1960) and pupillary hypofunction in response to pain, light, or exercise (May, 1948). Astrup (1962) found inadequate vascular responses to cold in all types of schizophrenia, and Hall and Stride (1954) reported higher thresholds to thermal pain. Reduced GSR responses have been noted to Hock, Kubis, and Ronke (1944), Jurko et al. (1952) and Solomon et al. (1939).

Except for the GSR, hyporeactivity in chronic patients may be partly a function of their higher basal levels. The "law of initial values" (Lacey, 1956; Wilder, 1950) predicts reduced responses when psychological systems approach homeostatic limits. Williams' (1953) and Reynolds' (1962) data are consistent with this hypothesis. The latter author found that normals had the lowest basal levels and showed a typical increase in functioning under stress. Reactive schizophrenics displayed a reduced response but a higher base level; process patients yielded the highest initial values and showed the least change. Separating reactivity from base level will prove to be even more difficult if, as Shakow (1963) suggests, the rate of adaptation to stimuli is significantly slower in schizophrenics than normals.

Reactivity is often less in chronic than acute schizophrenics, and lower in process than reactive patients. Malmo and Shagass (1949) reported that early schizophrenics are hyperreactive and resemble anxiety neurotics in the EMG response to pain. Chronic patients showed a reduced muscular response to the same stimulus, although this was not true for heart rate (Malmo, et al., 1951). King (1958) found reactive schizophrenics hospitalized less than 8 weeks to be hyperreactive to mecholyl. Process schizophrenics proved to be hyporeactive. DeVault (1957) reported negative heart rate changes to his experimental stimuli in process patients, while reactive schizophrenics tended to show the same positive increases as normal controls.

Venables and Wing (1962) studied the relationship between arousal and ratings on a withdrawal scale (Venables & O'Connor, 1959) by schizophrenic patients' charge nurse. Arousal was measured by the two-flash threshold and the skin potential response. With the exception of those deluded patients who showed no incoherence of speech (essentially intact paranoid schizophrenics), increased physiological arousal was associated with increased ratings of social withdrawal.

The above findings are provocative but difficult to relate to other research. Most workers have measured resistance to an exosomatic current, a procedure not directly comparable with the recording of endosomatic potentials. Furthermore, the visual threshold measure has been infrequently used with psychiatric patients. In one study King (1962c) found no difference between schizophrenics and normals on the two-flash threshold, but his sample was small and distinctions within schizophrenics were not made. More recently Venables (1963b) repeated his results for withdrawal and found the two-flash threshold to be significantly related to the extent irrelevant information disturbed the card sorting behavior of nonparanoid schizophrenics. He also reported (Venables, 1963a) that, while normals are unaffected, the thresholds of schizophrenic patients are significantly altered by coincident noise. Further study of these phenomena is needed.

Funkenstein (1951) reported that amelioration of schizophrenic pathology is associated with a reduction in basal systolic blood pressure. Gellhorn (1953) and Gunderson (1953) also found that improvement in schizophrenia is correlated with a reduction in the ANS basal level of functioning. Weckowicz (1958) reported that reduced blood pressure response to the mecholyl test is associated with deteriorated size constancy. He suggested that the hyporeactivity in these patients is attributable to higher basal levels of sympathetic activity. An increase in GSR with clinical improvement has been noted by Solomon et al. (1939) and Hock et al. (1944). Finally, Reynolds (1962) found that polygraph profiles of schizophrenics receiving tranquilizing medication were more like those of normals than those of patients off medication at the time of his experiment.[6]

Reactivity and the properties of the stimulus. Paintal (1951) reported that psychotics gave a reduced GSR response to threat of pain. Ray (1963) failed to confirm this finding in a study of female schizophrenics, but he obtained significant set differences. With instructions simply to listen to a list of "loaded" and "neutral" words, no difference in GSR between schizophrenics and normals was noted; with instructions to respond with an association, however, the normal GSR increased considerably more than that of the schizophrenics. This was true regardless of whether the schizophrenics proved to be "adequate" or "inadequate" verbal responders. Furthermore, while the GSRs of normals were greatest to the loaded words under both conditions, the schizophrenics yielded differential responses only when actually responding.

Venables (1960) found that the GSR of schizophrenic subjects varies with the stimulus context in a manner similar to RT. Schizophrenic and normal subjects were presented discrete visual and auditory stimuli, with or without a continuous, collateral stimulus of the opposite modality. The experimental conditions did not significantly affect the GSR of normal subjects. A low illumination and quiet condition was compared with a bright illumination and noisy condition; in "active schizophrenics" the former yielded more GSRs which were of greater intensity and shorter latency. In withdrawn schizophrenics the bright, noisy condition produced the same number of GSRs as the quiet condition, but latency was shorter. Thus, as with RT, additional stimuli appear to modify the speed of response, with the most intense stimuli affecting withdrawn schizophrenics the most.

---

[6] Tourney et al. (1962) reported that chronic schizophrenics show a disturbance in the mechanism concerned with the transformation of chemical to kinetic energy which seems to parallel the findings cited in this section. The specific activity of chemicals involved in carbohydrate breakdown is greater under basal conditions in chronic schizophrenic patients than normal subjects. However, the mobilization of these products in response to stimulation was significantly less in those patients than in normal controls.

Venables recalled that any sensory input may serve both a cue and a nonspecific arousal function (Hebb, 1955) and that this latter dimension may describe an inverted U function (Malmo, 1958). He suggested that with moderate stimulation active schizophrenics function at an optimal level. With increased intensity there is a paradoxical depression of response. On the other hand, moderate stimuli are less effective for withdrawn patients, and increases in stimulus intensity improve performance in a linear fashion. Unfortunately, no information was provided about basal potentials under both conditions of stimulation, rendering the results and the interpretation inconclusive.

Leach (1960) studied ocular nystagmus in response to rotation at various acceleration speeds. He reported that

> with increasing intensity of the stimulus schizophrenic deficit decreased. This tendency is demonstrated most clearly in latency of nystagmus to onset rotation. Schizophrenic deficit was less than one-fourth for moderate and strong intensities than for a mild intensity [p. 308].

The response of normals remained relatively constant in relation to changes in stimulus intensity. These findings parallel those of Venables and Tizard (1958), Lang (1959), and King (1962b), who reported a decrease in relative psychomotor deficit with increases in the physical intensity of the stimulus. The fact that the semicircular canals regulate muscle tonus in the body lends added meaning to this parallel. Furthermore, Angyal and Blackman (1940) showed an association within schizophrenia between nystagmic deficit and disturbance in muscular tonus.

Reduced caloric and rotational nystagmus has also been reported in childhood schizophrenia. Colbert, Koegler, and Markham (1959) found that all the subjects in the sample for whom nystagmus was absent carried the schizophrenic diagnosis. When nystagmus was present, the shorter its duration, the greater the probability of schizophrenia. It is interesting to note that, as with psychomotor behavior (King, 1954, 1961), intraindividual variability was as pathognomic for schizophrenia as the reduced response. These authors ruled out the possibility that a lesion in the major vestibular pathways could be the causal agent because none of the usual collateral symptoms were observed. They speculated that vestibular activity may be inhibited at the level of the caudal midbrain, in a manner neurologically analogous to the inhibition of auditory attention demonstrated by Hernandez-Péon and Associates (1956).

Another approach suggests that both vestibular and psychomotor deficit are attributable to a disturbance of proprioception. Rosenbaum et al. (1959) using shock and no shock conditions studied simple RT of chronic schizophrenics and normal subjects who were administered either LSD-25, amobarbital sodium, or phencyclidine hydrochloride (sernyl). Under nonshock conditions, the sernyl normals and schizophrenics had significantly slower latencies than the other two drug groups. With shock, both the schizophrenics and the sernyl subjects reduced their scores to the level of the other subjects. Similar parallels between schizophrenic and sernyl normals were noted in rotary pursuit learning and weight discrimination. Unlike LSD-25, sernyl seems to produce primary, rather than secondary, schizophrenic symptoms. Furthermore, when it is administered to schizophrenic patients, pathology is markedly exacerbated, and this condition may last for several weeks (Luby, Gottlieb, Cohen, Rosenbaum, & Domino, 1962).

In normal subjects sernyl depresses "central integrating mechanisms involving various sensory modalities such as touch, pain, and proprioception [Luby et al., 1962, p. 64]"; subjects report alterations of body image and feelings of estrangement and unreality, and they display progressive disorganization of thought, inability to maintain set, and impairment of abstract thinking. The drug effects show a further parallel to schizophrenia in that increased levels of respiratory and cardiovascular activity have been reported (Meyer, Greifenstein, & Devault, 1959). Cohen, Rosenbaum, Luby, & Gottlieb (1962) studied the effects of schizophrenia, sernyl, LSD, and amytal on proverb interpretation and a serial sevens task. Under nondrug conditions, the normal groups were superior to the schizophrenic patients. LSD and amytal produced insignificant decrements in the performance of the normals. However, sernyl subjects approximated the level of the patients. Cohen et al. (1962) proposed that both sernyl normals and schizophrenics suffer from a basic proprioceptive deficit, this failure of feedback accounting for both psychomotor and cognitive disturbance.

> The *S* is impaired in his ability to provide himself with those response-produced cues which normally function to enhance stimulus discrimination and relevant response selection [p. 84].

This formulation meaningfully parallels the hypothesized breakdown in self-instruction previously raised to account for the findings of verbal reinforcement studies (Buss & Lang, 1965).

The fact that schizophrenics display increased deficit when many competing stimuli are present has already been amply documented. Recently a number of researchers have reported a decrease in behavior disturbance under conditions of decreased stimulation Cohen, Rosenbaum, Dobe, and Gottlieb (1959), Harris (1959), and Smith, Thakurdas, and Lawes (1961) all found that schizophrenics who have experienced sensory deprivation show less discomfort and some improvement under these conditions. Reitman and Cleveland (1964) reported that schizophrenics showed an increase in tactile sensitivity and gave more accurate estimates of body size, following sensory deprivation. These positive effects were not observed in similarly treated normal subjects.

Cohen, Luby, Rosenbaum, and Gottlieb (1960) and Lawes (1963) studied the effects of sensory deprivation on normal subjects who had been administered sernyl. Like schizophrenics, sernyl normals evidenced less disorder of thought, attention, and perception under deprivation conditions than under normal conditions of stimulation. Lawes suggested that the schizophrenic is overwhelmed by normal levels of external stimulation and incapable of managing sensory inputs. Callaway and Dembo (1958) described a narrowing of attention which they related to a kind of crowding out of meaningful stimuli by the high level of central sympathetic activity. These authors also reported disturbance of size constancy, reduced EMG, GSR reactivity, and learning deficit produced by drugs that initiate widespread sympathetic activity (nerve gas, amylnitrite, and amphetamine).

There are few studies of cortical responsitivity in schizophrenia that can be meaningfully related to psychological deficit. EEG abnormalities have frequently been reported (Ellingson, 1954), but differences have too often been judged rather than measured. Gromoll (1961) tested the hypothesis that reactive schizophrenics would be more responsive and show higher levels of cortical arousal than process patients or controls. No significant difference between groups were obtained on such measures as alpha blocking. However, the author reported that process subjects, rather than reactive schizophrenics, tended to maintain the highest activation levels.

Shagass and Schwartz (1961) have developed techniques for studying evoked responses in the somatosensory cortex of man. Subjects receive electric shock to the wrist, and potentials are recorded from EEG surface electrodes. By repeating the stimulation and averaging across trials, the form of evoked potentials may be determined. In a number of experiments a trial has consisted of two shocks separated by an interval varying in milliseconds. Normal subjects, neurotic depressives, and anxiety patients showed a response to the second shock at around 20 milliseconds that equalled the initial evoked potential. Psychotic depressives and schizophrenics showed a much reduced response in this early recovery phase (Shagass & Schwartz, 1961, 1962). Data consistent with these findings were reported by Purpura, Pool, Ransohoff, Freeman, and Houspian (1957), who found that the direct stimulation of the exposed human cortex yielded much later recovery in two schizophrenic patients than in the nondiseased cortex of a patient with a tumor.

Shagass and Schwartz (1963) also studied the effects of shock intensity on evoked cortical potentials in different psychiatric patients. They reported that the intensity response gradients of a mixed group of patients (including schizophrenics and psychotic depressives) were steeper than those of normal subjects or dysthymic patients. These data parallel the intensity gradients obtained for RT and suggest that studies relating cortical potentials and psychomotor behavior may be of considerable value in studying schizophrenia.

Summary

The picture of schizophrenic deficit that emerges from these findings is remarkably consistent across a number of very different response systems. Latency and/or amplitude of psychomotor, vestibular, cardiovascular, sweat gland, and cortical EEG responses are reduced, relative to normal subjects. In at least three of the above systems and in verbal association, excessive intraindividual variability of response has also proved to be pathognomic of schizophrenic disorder. In addition, the levels of cardiovascular activity and muscular tension are unusually high among these patients. All these behaviors—reduced responsivity, deterioration of associational or psychomotor control, and high somatic tension—are positively related to increased withdrawal or clinically judged exacerbation of the illness. They are more marked for chronic and process schizophrenics than for acute and reactive patients. These relationships do not appear to hold for relatively intact paranoids, and perhaps not for early schizophrenics (recent, first admissions).

The experimental manipulation of stimulus intensity has yielded consistent data in studies of RT, ocular nystagmus, and cortical potentials. Deficit in chronic schizophrenia is greatest for low-intensity inputs and least when stimulus amplitude is high. Related to these findings are results of distractibility experiments, which reveal both the schizophrenic's susceptibility to irrelevant cues and his improved performance when background noise is reduced.

A host of studies indicate set disturbances in schizophrenia. On the one hand, schizophrenics are unable to maintain response readiness, and response latency increases if stimuli are presented in more than one modality. On the other hand, these patients seem unduly influenced by a previous set, and responses persist long after they are demonstrably ineffective.

In general, the hypothesis that schizophrenic deficit is attributable to the interference of competing stimuli, internal or external, receives considerable support. The theory that schizophrenics are underaroused may be maintained only if studies of activity level are ignored. This latter research argues that even long-term chronic patients may be physiologically hyperaroused relative to normal subjects, although the frequency and amplitude of overt behavior is greatly reduced.

DISCUSSION

Theory

In the previous review (Buss & Lang, 1965) three general theories of schizophrenic deficit were considered. The first of these can be roughly described as social or interpersonal in emphasis: deficit is variously attributed to social censure, oversensitivity to affective stimuli, or lowered social motivation. The second approach holds that schizophrenic deficit is a consequence of regression. A third view argues that associative interference underlies many instances of the schizophrenic's behavior disturbance.

Regression theory received little support from data summarized in the first paper, and the current review adds nothing that alters conclusions drawn there. In this discussion, three motivational constructs (social motivation, drive, and arousal) and a more broadly conceived interference theory will be considered as explanations of schizophrenic deficit.

Social motivation. The hypothesis that schizophrenics suffer from lowered social motivation was examined in the previous review. It was seen that schizophrenics and normals respond similarly to general encouragement or chastisement on laboratory tasks. Furthermore, when specific responses are reinforced, differences between groups may be attributed to the greater value of information about performance for the schizophrenic subject. A guidance function is served for patients, which normals apparently provide for themselves. For example, punishment breaks up the perseverative behavior of psychotics, resulting in a closer approximation of normal performance.

The fact that schizophrenics improve more than normals when aversive, physical reinforcers are used has been interpreted to mean that schizophrenics' response to social reward is reduced. However, the research reviewed in the present paper suggests that the intense stimuli employed in these experiments serve to emphasize relevant cues and focus attention, rather than function as special motivators for an indifferent patient.

Some theorists argue that the schizophrenic's problem is not *undersensitivity* to social motivators; it is held that their *oversensitivity* to the affective meaning of stimuli disrupts performance. While affective stimuli may increase deficit, this property is not restricted to one type, such as social censure. In fact, the considerable variety of stimuli (symbolic, human versus nonhuman, etc.) capable of producing these effects calls into questions the value of a category so loosely defined. In the previous paper, the authors suggested that the deficit produced by so-called affective or emotionally arousing stimuli is due to an inability to inhibit irrelevant associations. Most of these stimuli instigate more associations than do the neutral comparison stimuli. Evidence presented here indicates that the capacity to suppress *any* intruding cognition is greatly reduced in schizophrenia.

In summary, the hypotheses that schizophrenics are indifferent to social stimuli or particularly sensitive to the affective meaning of stimuli, have very limited value. Experiments relevant to both views are more parsimoniously interpreted in the context of interference theory, which will be reconsidered after a discussion of drive and arousal.

Drive. Negative drive theory applies Pavlov's notion of protective inhibition in schizophrenia to the learning process. Specifically, schizophrenics are held to develop reactive inhibition faster than normals and should therefore show greater reminiscence. Concerning reactive inhibition, there is no consistent evidence that schizophrenics classically condition slower than normals. Concerning reminiscence, the results are similar: no established difference between schizophrenics and normals. It seems safe to conclude that negative drive theory is incorrect.

Mednick, labeling the potential schizophrenic as high-anxious, used the Spence-Taylor approach in making predictions: faster conditioning in simple situations, slower conditioning in complex situations, and flattened generalization gradients. As we showed earlier, these predictions have received only weak support, and there is strong opposing evidence. Thus Mednick's theory has, in general, not been sustained by research findings.

What appears to be wrong with the theory is its specification of anxiety as the crucial drive that leads to schizophrenia. While it is true that many schizophrenics appear anxious, this could as readily be a reaction to incapacity as a cause of it. The theory is embarrassed not only because the predictions from anxiety theory are not supported but also because more chronic and severe schizophrenics show less clinical anxiety. The fact that chronic, withdrawn patients frequently have high-somatic activity levels appears partially to save the theory. However, Mednick has already explained that the chronic schizophrenic's associational defense successfully eliminates anxiety!

These weaknesses of Mednick's theory do not necessarily apply to all drive theories. In fact, drive theory can be shown to be consistent with much research evidence if it is assumed that: (*a*) it is a generalized drive state rather than a specific one such as anxiety, and (*b*) generalized drive can be measured by, or is the same as, physiological arousal. Two sets of facts seem to fit a generalized drive theory. First, schizophrenics tend to be over-aroused, the physiological hyperactivity varying directly with chronicity and/or severity. Second, schizophrenic deficit also varies directly with chronicity and/or severity. These facts can be combined in a causal sequence: schizophrenic deficit is due to the disruptive effects of an excessively high arousal or generalized drive state. Stated this way drive theory can be seen to be one variant of interference theory.

Arousal. Complementary to drive theory is what may be called arousal theory. This view was originally based on the neurological speculations of Hebb (1955) and Lindsley (1951), the EEG work of the latter, and studies of the ANS and muscle tension system by Freeman (1948), Duffy (1962) and Malmo (1958). This conception orders behavior on a continuum from deep sleep to intense excitement.

These behavioral states are held to be a function of the degree of diffuse activity in the lower brain, particularly in the reticular formation. From this site collateral impulses ascend to the cortex and descend to the ANS. Alertness, attention, and reactivity are thus determined by the organism's level of "arousal." As with the social motivational point of view, schizophrenics have been held to be both overaroused and underaroused.

The hypothesis that schizophrenics suffer from an underactive arousal mechanism would seem to receive support from studies demonstrating psychomotor and physiological hyporeactivity in chronic patients. However, Malmo (1958) argues cogently that activation is measured more directly in studies of basal physiological level than in research on responsivity. Thus, studies showing high resting somatic activity in schizophrenia would indicate that schizophrenics are generally hyperaroused rather than the opposite. Furthermore, their reduced responsivity is not inconsistent with this view. Malmo (1958), Lacey (1956), and Wilder (1950) have all presented evidence that responsivity progressively decreases when plotted on an abscissa of increasing activation (defined by base activity level). A similar function is obtained in normal subjects when "adequacy of performance" in a complex psychological task replaces responsivity on the ordinate. These facts not only argue that testable chronic schizophrenics are habitually in a hyperaroused state, but in this context the performance deterioration of schizophrenics appears to be analogous to the psychological stress response of a normal subject. However, the symmetry of this analogy is only apparent. Whereas it is complex functioning of normal subjects that mainly suffers under stress (while perhaps more primitive and less adequate but well organized responses emerge), the schizophrenic patient shows deterioration of the simplest and most fundamental behaviors. For example, the schizophrenic performs poorly on a RT task, not because he is anxious, an overready impulsive responder, but because the stimulus seems to arrive unexpectedly. He is not prepared or set, and the response is slow and reduced in amplitude. The psychomotor performance of chronic schizophrenics is more similar to that of aged normals or young adults with general cerebral damage than to that of psychologically stressed normals or anxiety neurotics.

In summary, the underarousal theory of schizophrenia, in terms of the nonspecific projection system, is directly contradicted by most of the psychophysiological research reviewed here, and it may be considered incorrect. The hypothesis that schizophrenics are overaroused receives some support. However, the exact mechanism by which overarousal can produce hyporesponsivity, high-response variability, inattention, disturbances of set and association, and the other symptoms of chronic schizophrenia is yet to be explained.

Interference theory. Interference theory has focused mainly on association and attention-set. The associations of schizophrenics are idiosyncratic and deviant, and they deteriorate performance because they serve as distractors. Schizophrenics have difficulty in focusing on relevant stimuli and excluding irrelevant stimuli, in maintaining a set over time, in shifting a set when it is necessary, in instructing themselves and in pacing themselves, and generally in performing efficiently, in Wishner's sense (1955). These difficulties are pervasive, occurring over a wide range of perceptual, motor, and cognitive tasks. In brief, interference theory, as a broad explanation of schizophrenic deficit, has clearly been supported by research findings and appears to be the only theory comprehensive enough to account for what is known.

The generality of interfering effects suggests a fundamental sensori-motor defect. However, the reactions of patients vary somewhat according to subtype. The defect is seen most clearly in the behavior of chronic, withdrawn patients; acute schizophrenics and particularly early paranoids seem to be compensating for their disability. In many tasks they are overprecise or overresponsive. The fact that in some experiments, among paranoid schizophrenics only the most chronic cases show deficit, suggests that their bizarre attempts at organizing the world may have functional value. Support is gained for Bleuler's contention that many of these behaviors are secondary symptoms—responses to the fundamental disturbance, rather than intrinsic expressions of it.

The locus of the sensori-motor defect is a matter for speculation. It seems clear that the defect is not at the level of the peripheral sensors[7] or effectors, although feedback from the musculature and the ANS may contribute to the disturbance. Lacey and Lacey (1958a) suggest that attention, set, and psychomotor control are directly influenced by autonomic feedback. They propose the carotid sinus as one such steering mechanism: blood pressure changes stimulate the carotid, which has "a profound tonic and inhibitory effect" on cortical electrical activity, and thus alters the organism's orientation to the environment. Furthermore, these researchers have demonstrated a relationship between cardiac variability and failure to inhibit psychomotor responses. Recently, they have also shown that heart rate changes correlate with RT fore-period effects. This raises the interesting possibility that the motor and perceptual symptoms of schizophrenia are related to defects in this carotid-cortical mechanism.

The disturbance that appears in all studies of deficit concerns the initiation of responses to selected stimuli and the inhibition of inappropriate responses. All intelligent behavior represents a compromise between the demands of the immediate environment and a previously established set of the organism, but the schizophrenic makes a uniquely poor bargain. External stimuli, associational and biological "noise," routinely suppressed by normal subjects, intrude, and responses to the appropriate stimuli are not made.

These facts suggest that researchers in schizophrenia should concentrate on the processes by which stimuli adapt out or habituate and response competition is resolved. The ascending reticular activating system is the neurological site of greatest relevance. In addition to general arousal, this system appears to have a specific alerting or focusing function. Hernandes-Péon and his associates (1956) demonstrated that cortical potentials in the cochlear nucleus of the cat, normally elicited by a tone, were suppressed when a competing odor of fish or a jar of mice was simultaneously presented. These authors write:

> Attention involves the selective awareness of certain sensory messages with the simultaneous suppression of others. . . . During the attentive state, it seems as though the brain integrates for consciousness only a limited amount of sensory information, specifically, those impulses concerned with the object of attention [p. 332].

The data on schizophrenic deficit are consistent with the hypothesis that such sensory inhibition centers are defective. These centers and the related behavior should be given extensive study in schizophrenic patients.

Methodological Considerations

It is appropriate that a research review should conclude on a methodological note. The issues raised are many, and their listing amounts to a set of guidelines and admonishments to future investigators.

1. A number of studies have shown that schizophrenics as a group are more variable than normals, and no one regards the nonpsychiatric population as being especially homogeneous. Furthermore, schizophrenics are known to vary in the extent of deficit in relation to several variables which are usually dichotomized: mild-severe, acute-chronic, reactive-process, good premorbid-poor premorbid, and paranoid-nonparanoid. It seems likely that these variables overlap, but empirical data are limited. We need studies relating these dimensions of schizophrenia to each other, as well as more precise data on their relation to deficit.

The paranoid-nonparanoid dichotomy is of special interest. For over a century there has been doubt about including paranoids under the heading of schizophrenia or keeping them separate as "paranoid conditions." Paranoids have been found to show less deficit (e.g., Payne & Hewlitt, 1960), and clinically they have been observed to show less thought disorder and less deterioration over time than have schizophrenics of other subgroups. However, these statements are not true of all paranoids; some patients with delusions do manifest considerable deficit and deterioration of thought processes. Perhaps the presence of delusions is less important than the relative absence of deficit. Stated another way, perhaps the important dimension is intactness of sensori-motor and intellectual processes, and the paranoid-nonparanoid distinction partially reflects or is partially correlated with this dimension.

---

[7] Schizophrenics are no different from normals in pure tone threshold and speech reception but are disrupted more quickly and easily by auditory feedback and noise (Ludwig, Wood, & Downs, 1962).

Recently Johannsen and his associates (1963) examined correlations between different measures used to describe schizophrenics. High correlations were found between placement on process-reactive, acute-chronic, and good-poor premorbid scales. Only the paranoid-nonparanoid dimension appeared to be an independent dimension. Furthermore, this latter dichotomy was the only one that yielded a significant difference on a double alternation learning task. Whether delusional behavior is an epiphenomenon in low-deficit schizophrenics or a positive effort to reduce deficit as was suggested earlier, future investigators must consider paranoid symptoms in selecting experimental samples.

A less known source of variability among schizophrenics may be found in sex differences. The subjects in most research have been men, with a minority of experiments including both sexes or using women only. It is possible that results found with men cannot be generalized to women, and sex differences might account for some of the discrepancies in results that occur among studies otherwise comparable.

The importance of this issue is pointed up by Schooler's (1963) study of affiliation. He found that the relationships that held for men did not hold for women, and vice versa, which led him to conclude:

> A major implication of the study is that theories based on experimental findings with one sex cannot be generalized to explain the behavior of chronic schizophrenics of the other sex [p. 445].

2. In many instances the range of tasks used to study deficit is not sufficient to sustain the broad conclusions of the investigators. For example, on the basis of demonstrated deficiency on conceptual tasks, some researchers have concluded that the basic problem of schizophrenic deficit is an inability to handle concepts. Taken at face value, this conclusion is an overgeneralization because of the absence of evidence that schizophrenics show no deficit on nonconceptual tasks. In the light of the evidence with nonconceptual tasks, the conclusion is patently false. Generalizations about schizophrenic deficit require a sampling of tasks that tap a variety of psychological functions.

It would be of considerable help if we knew more about what various tasks are measuring and their relations to each other. The appropriate tool is factor analysis, which has been employed mainly by English researchers such as Payne and his collaborators (Payne, Mattussek, & George, 1959; Payne & Hewlitt, 1960).

3. When the investigator is interested in particular characteristics of his stimuli, a special problem may arise. He may assume, for example, that some of his stimuli are "affect-laden" without having any evidence for this assertion. A priori statements that stimuli differ along a dimension such as "emotionally arousing" cannot be accepted. It behooves the investigator to present evidence on this point, and the evidence must be independent of the effects obtained with his dependent variable. A similar problem appeared in studies of positive and negative incentives. These experiments were generally interpreted in a social-motivational context, while the more important differences in the degree of information conveyed by these stimuli were largely ignored.

4. General methodological problems in psychophysiological research have been adequately described elsewhere (Lacey, 1956; Lacey & Lacey, 1958b). However, these difficulties are accentuated when schizophrenics are the experimental subjects. For example, the low-positive correlations between physiological measures noted in studies of normal subjects may be lower or even negative in schizophrenics. Single measures of arousal or drive are necessarily misleading. Thus, hypoactivity in schizophrenia is frequently found for skin resistance, while muscle tension is generally reported to be high. Such results are provocative, and further study of relationships between sweat gland, cardiovascular, and muscle tension systems may prove valuable.

The pervasive use of drugs in the treatment of psychosis creates problems for both the behavioral and psychophysiological investigator.

For example, Reynolds (1962) found a significant interaction between diagnostic subtype (process-reactive) and tranquilizer-nontranquilizer conditions in a study of somatic responses in schizophrenia. No research should be undertaken unless the drug variable is properly controlled.

Researchers have begun to emphasize individual variability in behavioral studies of schizophrenia. Investigations of somatic inter- and intrasession variability are also needed. Furthermore, there may be profit in studying somatic responses recorded concurrently with tasks that elicit deficit. Lacey and Lacey (1958a) have reported important relationships between autonomic activity and psychomotor functioning in normals, and studies cited here encourage this experimental strategy.

Better estimates are needed of resting somatic activity levels. Despite the elaborate care of some investigators, what purport to be differences in basal levels between psychotics and normals may actually be differences in reaction to the laboratory situation. Long-term studies are needed in which information is telemetered from patients while they proceed with the usual hospital routine.

5. While only psychological deficit in schizophrenia has been considered in this review, it is important to reaffirm that these patients share many of the characteristics of deficit with other psychiatric disorders and cases of cerebral damage. The psychomotor retardation, inattention, increased response variability, muscle tension and ANS hyperactivity, and even to some extent the associative disturbance, may be found in many aged, paretic, severe epileptic, or arterial sclerotic patients. Deficit behavior can be produced in normal subjects through the administration of drugs or surgical intervention, and there is some evidence that it may be manipulated by brain stimulation (Heath, 1954).

There is ample evidence that severity of psychopathology and psychological deficit are positively related. Some theorists hold that this is the only meaningful relationship between deficit and diagnosis, and they argue that specific consideration of schizophrenia is superfluous. They emphasize the unity of deficit in psychiatric illness and suggest a common neurological defect underlies all its manifestations.

Certainly, further demonstrations that schizophrenics differ from normals are not needed. If the schizophrenic label has experimental validity, the deficit specific to this diagnosis must be more clearly defined. Are variables such as maternal censure or pictures of hospital aids uniquely important to the behavior of schizophrenics, or might they similarly influence the responses of other patient groups? An answer to this question can only come from studies employing control subjects other than normals, i.e., anxiety neurotics, aged, epileptic, brain-damaged or other chronically ill patients. While some experiments have compared schizophrenics to these groups, the evidence is fragmentary and the interpretations usually emphasize the safer, more reliable distinction between normality and psychosis.

The theoretical point of studies in this area often needs sharpening. Deficit is simply performance decrement. In trying to explain it we must distinguish between what is basic to the disorder and what is epiphenomenal. For example: Is the schizophrenic's anxiety the instigator of deficit, or is it an individual reaction to an insidious and pervasive sensorimotor defect? Issues of this type will tax the ingenuity of the behavioral researcher.

In summary, the problem of psychological deficit remains as broad and as challenging now as in 1944. However, the last 20 years have done much to clarify fundamental symptoms and define conditions which increase or decrease deficit. Many theories have failed to receive empirical support and may now be discarded. Fruitful lines of investigation have also been revealed, and the researcher today, guided by this work, is better equipped to discover the basic nature of schizophrenia.

REFERENCES

ACKER, C. W. An investigation of the variability in repeated psychophysiological measurements in tranquilized mental patients. *American Psychologist*, 1963, **18**, 454. (Abstract)

ALTSCHULE, M. D., & SULZBACH, W. M. Effect of carbon dioxide on acrocyanosis in schizophrenia. *Archives of Neurology and Psychiatry*, 1949, **61**, 44–55.

Angyal, A. The perceptual basis of somatic delusions in a case of schizophrenia. *Archives of Neurology and Psychiatry,* 1935, **34**, 270–279.

Angyal, A. The experience of the body-self in schizophrenia. *Archives of Neurology and Psychiatry,* 1936, **35**, 1029–1053.

Angyal, A., & Blackman, N. Vestibular reactivity in schizophrenia. *Archives of Neurology and Psychiatry,* 1940, **44**, 611–620.

Angyal, A., & Blackman, N. Paradoxical vestibular reactivity in schizophrenia under influence of alcohol, of hyperpnea and CO_2 inhalation. *American Journal of Psychiatry,* 1941, **97**, 894–903.

Angyal, A., & Sherman, N. A. Postural reactions to vestibular stimulation in schizophrenics and normal subjects. *American Journal of Psychiatry,* 1942, **98**, 857–862.

Arieti, S. *Interpretation of schizophrenia.* New York: Bruner, 1955.

Astrup, C. *Schizophrenia: Conditional reflex studies.* Springfield, Ill.: Thomas, 1962.

Barratt, E. S. CNS correlates of intra-individual variability of ANS activity. Paper read at American Psychological Association, St. Louis, September 1962.

Bender, Lauretta, & Schilder, P. Unconditioned and conditioned reactions to pain in schizophrenia. *American Journal of Psychiatry,* 1930, **87**, 365–384.

Blaufarb, H. A demonstration of verbal abstracting ability in chronic schizophrenics under enriched stimulus and instructional conditions. *Journal of Consulting Psychology,* 1962, **26**, 471–475.

Bleke, R. C. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *Journal of Personality,* 1955, **23**, 479–488.

Bleuler, E. *Dementia praecox or the group of schizophrenias.* New York: International University Press, 1950.

Buck, C. W., Carscallen, H. B., & Hobbs, G. E. Temperature regulation in schizophrenia: I. Comparison of schizophrenic and normal subjects. II. Analysis by duration of psychosis. *Archives of Neurology and Psychiatry,* 1950, **64**, 828–842.

Buss, A. H., & Lang, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *Journal of Abnormal Psychology,* 1965, **70**, 2–24.

Calloway, E., III, & Dembo, D. Narrowed attention. *Archives of Neurology and Psychiatry,* 1958, **79**, 74–90.

Campbell, D. A study of some sensorimotor functions in psychiatric patients. Unpublished doctoral dissertation, University of London, 1957.

Carrigan, P. M. Selective variability in schizophrenia. *American Psychologist,* 1963, **18**, 427. (Abstract)

Chapman, L. J. Distractibility in the conceptual performance of schizophrenics. *Journal of Abnormal and Social Psychology,* 1956, **53**, 286–291.

Chapman, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *Journal of Abnormal and Social Psychology,* 1958, **56**, 374–379.

Chapman, L. J. Stimulus generalization and verbal behavior in schizophrenia. Paper read at Midwestern Psychological Association, Detroit, May 1962.

Chapman, J., & McGhie, A. A comparative study of disordered attention in schizophrenia. *Journal of Mental Science,* 1962, **108**, 487–500.

Claridge, G. The excitation-inhibition balance in neurotics. In H. J. Eysenck (Ed.), *Experiments in Personality.* Vol. 2. London: Routledge & Kegan Paul, 1960. Pp. 107–154.

Cohen, B. D., Ludy, E. E., Rosenbaum, G., & Gottlieb, J. S. Combined sernyl and sensory deprivation. *Comprehensive Psychiatry,* 1960, **1**, 345–348.

Cohen, B. D., Rosenbaum, G., Dobe, Shirley I., & Gottlieb, J. S. Sensory isolation: Hallucinogenic effects of a brief procedure. *Journal of Nervous and Mental Diseases,* 1959, **129**, 486–491.

Cohen, B. D., Rosenbaum G., Luby, E. D., & Gottlieb, J. S. Comparison of phencyclidine hydrochloride (sernyl) with other drugs. *Archives of General Psychiatry,* 1962, **6**, 395–401.

Cohen, B. D., Senf, Rita, & Huston, P. E. Perceptual accuracy in schizophrenia, depression, and neurosis, and effects of amytal. *Journal of Abnormal and Social Psychology,* 1956, **52**, 363–367.

Cohen, M. Ocular findings in 323 patients with schizophrenia: Preliminary report. *Archives of Opthalmology,* 1949, **41**, 697–700. (a)

Cohen, M. Preliminary report on ocular findings in 323 schizophrenic patients. *Psychiatric Quarterly,* 1949, **23**, 667–671. (b)

Colbert, E. G., Koegler, R. R., & Markham, C. H. Vestibular dysfunction in childhood schizophrenia. *Archives of General Psychiatry,* 1959, **1**, 600–617.

Cooper, R. Objective measures of perception in schizophrenics and normals. *Journal of Consulting Psychology,* 1960, **24**, 209–214.

Crookes, T. G. Size constancy and literalness in the Rorschach test. *British Journal of Medical Psychology,* 1957, **30**, 99–106.

Crumpton, Evelyn. Persistence of maladaptive responses in schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **66**, 615–618.

DeVault, S. Physiological responsiveness in reactive and process schizophrenia. *Dissertation Abstracts,* 1957, **17**, 1387.

Dillon, D. Differences between ascending and descending flicker-fusion thresholds among groups of hospitalized psychiatric patients and a group of normal control persons. *Journal of Psychology,* 1959, **48**, 255–262.

Donahoe, J. W., Curtin, Mary E., & Lipton, L. Interference effects with schizophrenic subjects in the acquisition and retention of verbal material. *Journal of Abnormal and Social Psychology,* 1961, **62**, 553–558.

Downing, R. W., Ebert, J. N., & Shubrooks, S. J. Effects of three types of verbal distractors on thinking in acute schizophrenia. *Perceptual and Motor Skills,* 1963, **17**, 881–882.

DRAGUNS, J. G. Responses to cognitive and perceptual ambiguity in chronic and acute schizophrenics. *Journal of Abnormal and Social Psychology,* 1963, **66**, 24–30.

DUFFY, ELIZABETH. *Activation and behavior.* New York: Wiley, 1962.

DUNN, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *Journal of Personality,* 1954, **23**, 48–64.

EDWARDS, A. S., & HARRIS, A. C. Laboratory measurements of deterioration and improvement among schizophrenics. *Journal of General Psychology,* 1953, **49**, 153–156.

ELLINGSON, R. J. Incidence of EEG abnormality among patients with mental disorders of apparently non-organic origin: Critical review. *American Journal of Psychiatry,* 1954, **111**, 263–275.

EYSENCK, H. J. Psychosis, drive, and inhibition: A theoretical and experimental account. *American Journal of Psychiatry,* 1961, **118**, 198–204.

EYSENCK, H. J., GRANGER, G. W., & BRENGELMANN, J. C. *Perceptual processes and mental illness.* New York: Basic Books, 1957.

FAIBISH, G. M. Schizophrenic response to words of mutiple meaning. *Journal of Personality,* 1961, **29**, 414–427.

FELDMAN, M. J., & DRASGOW, J. A visual-verbal test for schizophrenia. *Psychiatric Quarterly Supplement,* 1951, Part I, 1–10.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FRANKS, C. M. An experimental study of conditioning as related to mental abnormality. Unpublished doctoral dissertation, University of London, 1954.

FREEMAN, G. L. *The energetics of human behavior.* Ithaca: Cornell Univer. Press, 1948.

FREEMAN, G. L., & PATHMAN, J. H. Physiological reactions of psychotics to experimentally induced displacement. *American Journal of Psychiatry,* 1943, **100**, 406–412.

FREEMAN, H., & RODNICK, E. H. Autonomic and respiratory responses of schizophrenics and normal subjects to changes of intrapulmonary atmosphere. *Psychosomatic Medicine,* 1940, **2**, 101–109.

FREEMAN, H., & RODNICK, E. H. Effect of rotation on postural steadiness in normal and schizophrenic subjects. *Archives of Neurology and Psychiatry,* 1942, **48**, 47–53.

FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. Autonomic changes paralleling psychological changes in mentally ill patients. *Journal of Nervous and Mental Diseases,* 1951, **114**, 1–18.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality,* 1952, **21**, 253–276.

GELLHORN, E. *Physiological foundations of neurology and psychiatry.* Minneapolis: Univer. Minnesota Press, 1953.

GELLHORN, E. *Autonomic imbalance and the hypothalamus.* Minneapolis: Univer. Minnesota Press, 1957.

GINDIS, I. Z. The pathological changes in higher nervous activity in the various forms of schizophrenia. *Pavlov Journal of Higher Nervous Activity,* 1960, **10**, 434–439.

GOULD, L. N. Verbal hallucinations as automatic speech. The reactivation of dormant speech habits. *American Journal of Psychiatry,* 1950, **107**, 110–119.

GRISELL, J. L., & ROSENBAUM, G. Effects of auditory intensity on schizophrenic reaction time. *American Psychologist,* 1963, **18**, 394. (Abstract)

GROMOLL, H. F., JR. The process-reactive dimension of schizophrenia in relation to cortical activation and arousal. Unpublished doctoral dissertation, University of Illinois, 1961.

GUNDERSON, E. K. Autonomic balance in schizophrenia. Unpublished doctoral dissertation, University of California, Los Angeles, 1953.

HALL, K., & STRIDE, E. The varying responses to pain in psychiatric disorders: A study in abnormal psychology. *British Journal of Medical Psychology,* 1954, **27**, 48–60.

HARRIS, A. Sensory deprivation and schizophrenia. *Journal of Mental Science,* 1959, **105**, 235–237.

HEATH, R. (Ed.) *Studies in schizophrenia.* Cambridge: Harvard Univer. Press, 1954.

HEBB, D. O. Drives and the CNS (conceptual nervous system). *Psychological Review,* 1955, **62**, 243–254.

HENSCHEL, A., BROZEK, J., & KEYS, A. Indirect vasodilation in normal man and in schizophrenic patients. *Journal of Applied Physiology,* 1951, **4**, 340–344.

HERNANDEZ-PÉON, R., SCHERRER, H., & JOUVET, M. Modification of electrical activity in cochlear nucleus during "attention" in unanesthetized cats. *Science,* 1956, **123**, 331–332.

HIGGINS, J., & MEDNICK, S. A. Reminiscence and stage of illness in schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **66**, 314–317.

HOCK, P., KUBIS, J. F., & ROUKE, F. L. Psychogalvanometric investigations in psychoses and other abnormal states. *Psychosomatic Medicine,* 1944, **6**, 237–243.

HOSKINS, R. G. *The biology of schizophrenia.* New York: Norton, 1946.

HOWE, E. S. GSR conditioning in anxiety states, normals, and chronic functional schizophrenic subjects. *Journal of Abnormal and Social Psychology,* 1958, **56**, 183–189.

HUNT, J. McV., & COFER, C. Psychological deficit in schizophrenia. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944, Pp. 971–1032.

HUSTON, P. E., & SHAKOW, D. Learning in schizophrenia: I. Pursuit learning. *Journal of Personality,* 1948, **17**, 52–74.

HUSTON, P. E., & SHAKOW, D. Learning capacity in schizophrenia. *American Journal of Psychiatry,* 1949, **105**, 881–888.

Huston, P. E., Shakow, D., & Riggs, L. A. Studies of motor function in schizophrenia: II. Reaction time. *Journal of General Psychology,* 1937, **16,** 39–82.

Johannsen, W. J., Friedman, S. H., Leitschuh, T. H., & Ammons, Helen. A study of certain schizophrenic dimensions and their relationship to double alternation learning. *Journal of Consulting Psychology,* 1963, **27,** 375–382. (a)

Johannsen, W. J., Friedman, S. H., & Liccione, J. V. Visual perception as a function of chronicity in schizophrenia. *American Psychologist,* 1963, **18,** 364–365. (b) (Abstract)

Johnson, R. C., Weiss, R. L., & Zelhart, P. F. Similarities and differences between normal and psychotic subjects in response to verbal stimuli. *Journal of Abnormal and Social Psychology,* 1964, **68,** 221–226.

Jurko, M., Jost, H., & Hill, T. S. Pathology of the energy system: An experimental clinical study of physiological adaptiveness capacities in a nonpatient, a psychoneurotic, and an early paranoid schizophrenic group. *Journal of Psychology,* 1952, **33,** 183–189.

Karras, A. The effects of reinforcement and arousal on the psychomotor performance of chronic schizophrenics. *Journal of Abnormal and Social Psychology,* 1962, **65,** 104–111.

Kidd, Aline H. Monocular distance perception in schizophrenics. *Journal of Abnormal and Social Psychology,* 1964, **68,** 100–103.

King, G. F. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *Journal of Abnormal and Social Psychology,* 1958, **56,** 160–164.

King, H. E. *Psychomotor aspects of mental disease.* Cambridge: Harvard Univer. Press, 1954.

King, H. E. Some explorations in psychomotility. *Psychiatric Research Reports,* 1961, **14,** 62–86.

King, H. E. Anticipatory behavior: Temporal matching by normal and psychotic subjects. *Journal of Psychology,* 1962, **53,** 425–440. (a)

King, H. E. Reaction-time as a function of stimulus intensity among normal and psychotic subjects. *Journal of Psychology,* 1962, **54,** 299–307. (b)

King, H. E. Two flash and flicker fusion thresholds for normal and schizophrenic subjects. *Perceptual and Motor Skills,* 1962, **14,** 517–518. (c)

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Sciences,* 1956, **67,** 123–164.

Lacey, J. I., & Lacey, Beatrice C. The relationship of resting autonomic activity to motor impulsivity. *Research Publications of the Association for the Study of Nervous and Mental Diseases,* 1958, **36,** 144–209. (a)

Lacey, J. I., & Lacey, Beatrice C. Verification and extension of the principle of autonomic response stereotypy. *American Journal of Psychology,* 1958, **71,** 50–73. (b)

Lang, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *Journal of Abnormal and Social Psychology,* 1959, **59,** 263–268.

Lang, P. J., & Luoto, K. Mediation and associative facilitation in neurotic, psychotic, and normal subjects. *Journal of Abnormal and Social Psychology,* 1962, **64,** 113–120.

Lauro, L. P. Recall of nouns varying in clustering tendency by normals and schizophrenics. Unpublished doctoral dissertation, New York University, 1962.

Lawes, T. C. G. Schizophrenia, "sernyl" and sensory deprivation. *British Journal of Psychiatry,* 1963, **109,** 243–250.

Leach, W. W. Nystagmus: An integrative neural deficit in schizophrenia. *Journal of Abnormal and Social Psychology,* 1960, **60,** 305–309.

Leibowitz, H. W., & Pishkin, V. Perceptual size constancy in chronic schizophrenia. *Journal of Consulting Psychology,* 1961, **25,** 196–199.

Lester, J. R. Production of associative sequences in schizophrenia and chronic brain syndrome. *Journal of Abnormal and Social Psychology,* 1960, **60,** 225–233.

Lhamon, W. T., & Goldstone, S. The time sense. *Archives of Neurology and Psychiatry,* 1956, **76,** 625–629.

Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

Lovinger, E. Perceptual contact with reality. *Journal of Abnormal and Social Psychology,* 1956, **52,** 87–91.

Luby, E. D., Gottlieb, J. S., Cohen, B. D., Rosenbaum, G., & Domino, E. F. Model psychoses and schizophrenia. *American Journal of Psychiatry,* 1962, **119,** 61–67.

Ludwig, A. M., Wood, B. S., & Downs, M. P. Auditory studies in schizophrenia. *American Journal of Psychiatry,* 1962, **119,** 122–127.

Luria, A. R. *The nature of human conflicts.* New York: Liveright, 1932.

McDonough, J. M. Critical flicker frequency and the spiral aftereffect with process and reactive schizophrenia. *Journal of Consulting Psychology,* 1960, **24,** 150–155.

McGhie, A., & Chapman, J. Disorders of attention and perception in early schizophrenia. *British Journal of Medical Psychology,* 1961, **34,** 103–116.

McGinnies, E., & Adornetto, J. Perceptual defense in normal and schizophrenic observers. *Journal of Abnormal and Social Psychology,* 1952, **47,** 833–837.

Malmo, R. B. Experimental studies of mental patients under stress. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950, Pp. 229–265.

Malmo, R. B. Measurement of drive: An unsolved problem in psychology. In M. R. Jones (Ed.), *Nebraska Symposium on motivation.* Lincoln: Univer. Nebraska Press, 1958. Pp. 44–105.

Malmo, R. B., & Shagass, C. Physiological studies of reaction to stress in anxiety states and early schizophrenia. *Psychosomatic Medicine,* 1949, **11,** 9–24.

MALMO, R. B., & SHAGASS, C. Studies of blood pressure in psychiatric patients under stress. *Psychosomatic Medicine,* 1952, **14,** 82–93.

MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *Journal of Personality,* 1951, **19,** 359–375.

MALTZMAN, I., COHEN, S., & BELLONI, MARIGOLD. Associative behavior in normal and schizophrenic children. Technical report No. 11, 1963, University of California, Los Angeles, Contract Nonr 233 (50), Office of Naval Research.

MALTZMAN, I., SEYMORE, S., & LICHT, L. Verbal conditioning of common and uncommon word associations. *Psychological Reports,* 1962, **10,** 363–369.

MANDL, BILLE SUE T. An investigation of rigidity in paranoid schizophrenics as manifested in a perceptual task. *Dissertation Abstracts,* 1954, **14,** 2401–2402.

MARTIN, I. Levels of muscle activity in psychiatric patients. *Acta Psychologia,* Amsterdam, 1956, **12,** 326–341.

MAY, P. R. Pupillary abnormalities in schizophrenia and during muscular effort. *Journal of Mental Science,* 1948, **94,** 89–98.

MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *Journal of Abnormal and Social Psychology,* 1955, **51,** 536–542.

MEDNICK, S. A. Generalization as a function of manifest anxiety and adaptation to psychological experiments. *Journal of Consulting Psychology,* 1957, **21,** 491–494.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin,* 1958, **55,** 316–327.

MEYER, J. S., GRIEFENSTEIN, F., & DEVAULT, M. A new drug causing symptoms of sensory deprivation. *Journal of Nervous Mental Diseases,* 1959, **129,** 54–61.

MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychological Monographs,* 1953, **67**(20, Whole No. 370).

O'CONNER, N. Reminiscence and work decrement in catatonic and paranoid schizophrenics. *British Journal of Medical Psychology,* 1957, **30,** 188–193.

O'CONNOR, N., & RAWNSLEY, K. Two types of conditioning in psychotics and normals. *Journal of Abnormal and Social Psychology,* 1959, **58,** 157–161.

PAINTAL, A. S. A comparison of the GSR in normals and psychotics. *Journal of Experimental Psychology,* 1951, **41,** 425–428.

PAVLOV, I. P. *Conditioned reflexes and psychiatry.* Oxford: Oxford Univer. Press, 1941.

PAYNE, R. W., & HEWLITT, J. H. G. Thought disorder in psychotic patients. In H. J. Eysenck (Ed.), *Experiments in personality.* London: Routeledge & Kegan Paul, 1960, Volume II, Pp. 3–104.

PAYNE, R. W., MATTUSSEK, P., & GEORGE, E. I. An experimental study of schizophrenic thought disorder. *Journal of Mental Science,* 1959, **105,** 627–652.

PEARL, D., & BERG, P. S. D. Time perception and conflict arousal in schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **66,** 332–338.

PETERS, H. N., & MURPHREE, O. D. The conditioned reflex in the chronic schizophrenic. *Journal of Clinical Psychology,* 1954, **10,** 126–130.

PETURSSON, E. Electromyographic studies of muscular tension in psychiatric patients. *Comprehensive Psychiatry,* 1962, **3,** 29–36.

PFAFFMAN, C., & SCHLOSBERG, H. The conditioned knee jerk in psychotic and normal individuals. *Journal of Psychology,* 1936, **1,** 201–206.

PISHKIN, V., & HERSHISER, D. Respiration and GSR as functions of white sound in schizophrenia. *Journal of Consulting Psychology,* 1963, **27,** 330–337.

PISHKIN, V., SMITH, T. P., & LEIBOWITZ, H. W. The influence of symbolic stimulus value on perceived size in chronic schizophrenia. *Journal of Consulting Psychology,* 1962, **26,** 323–330.

PURPURA, D. P., POOL, J. L., RANSOHOFF, J., FREEMAN, H. J., & HOUSEPIAN, E. M. Observations on evoked dendritic potentials of the human cortex. *EEG and Clinical Neurophysiology,* 1957, **9,** 453–459.

RABIN, A. I. Time estimation of schizophrenics and non-psychotics. *Journal of Clinical Psychology,* 1957, **13,** 88–90.

RACHMAN, S. Disinhibition and the reminiscence effect in a motor learning task. *British Journal of Psychology,* 1962, **53,** 149–157.

RACHMAN, S. Inhibition and disinhibition in schizophrenics. *Archives of General Psychiatry,* 1963, **8,** 91–98.

RAUSH, H. L. Perceptual constancy in schizophrenia: 1. Size constancy. *Journal of Personality,* 1952, **21,** 176–187.

RAUSH, H. L. Object constancy in schizophrenia: The enhancement of symbolic objects and conceptual stability. *Journal of Abnormal and Social Psychology,* 1956, **52,** 231–234.

RAY, T. S. Electrodermal indications of levels of psychological disturbance in chronic schizophrenia. *American Psychologist,* 1963, **18,** 393. (Abstract)

REISMAN, M. N. Size constancy in schizophrenics and normals. Unpublished doctoral dissertation, University of Buffalo, 1961.

REITMAN, E. E., & CLEVELAND, S. E. Changes in body image following sensory deprivation in schizophrenic and control groups. *Journal of Abnormal and Social Psychology,* 1964, **68,** 168–176.

REYNOLDS, D. J. An investigation of the somatic response system in chronic schizophrenia. Unpublished doctoral dissertation, University of Pittsburgh, 1962.

REYNOLDS, G. A. Perceptual constancy in schizophrenics and "normals." *Dissertation Abstracts,* 1954, **14,** 1000–1001.

RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *American Journal of Psychiatry,* 1940, **97,** 214–225.

ROSENBAUM, G., COHEN, B. D., LUBY, E. D., GOTTLEIB, J. S., & YELEN, D. Comparison of sernyl with other drugs. *Archives of General Psychiatry,* 1959, **1,** 651–656.

ROSENTHAL, D., LAWLOR, W. G., ZAHN, T. P., & SHAKOW, D. The relationship of some aspects of mental set to degree of schizophrenic disorganization. *Journal of Personality,* 1960, **28,** 26–38.

RUBIN, L. Patterns of adrenergic-cholinergic imbalance in the functional psychoses. *Psychological Review,* 1962, **69,** 501–519.

SALZINGER, K. Shift in judgement of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. *Journal of Abnormal and Social Psychology,* 1957, **55,** 43–49.

SANDERS, R., & PACHT, A. R. Perceptual size constancy in known clinical groups. *Journal of Consulting Psychology,* 1952, **16,** 440–444.

SCHOOLER, C. Affiliation among schizophrenics: Preferred characteristics of the other. *Journal of Nervous and Mental Diseases,* 1963, **137,** 438–446.

SHAGASS, C., & SCHWARTZ, M. Reactivity cycle of somatosensory cortex in humans with and without psychiatric disorder. *Science,* 1961, **134,** 1757–1759.

SHAGASS, G., & SCHWARTZ, M. Some drug effects on evoked cerebral potentials in man. *Journal of Neuropsychiatry,* 1962, **3,** 49–58.

SHAGASS, C., & SCHWARTZ, M. Cerebral responsiveness in psychiatric patients. *Archives of General Psychiatry,* 1963, **8,** 177–189.

SHAKOW, D. Some psychological features of schizophrenia. In M. L. Reymert (Ed.), *Feelings and emotions.* New York: McGraw-Hill, 1950, Pp. 383–390.

SHAKOW, D. Segmental Set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry,* 1962, **6,** 17–33.

SHAKOW, D. Psychological deficit in schizophrenia. *Behavioral Science,* 1963, **8,** 275–305.

SHIPLEY, W. C. Studies of catatonia: VI. Further investigation of the perseverative tendency. *Psychiatric Quarterly,* 1934, **8,** 736–744.

SILVERMAN, J. Psychological deficit reduction in schizophrenia through response contingent noxious reinforcement. *Psychological Reports,* 1963, **13,** 187–210.

SMITH, S., THAKURDIS, H., & LAWES, T. G. G. Perceptual isolation and schizophrenia. *Journal of Mental Science,* 1961, **107,** 839–844.

SMITH, W. O. Rotary pursuit performance in reactive and process schizophrenics. Unpublished doctoral dissertation, Michigan State University, 1959.

SMOCK, C. D., & VANCINI, J. Dissipation rate of the effects of social censure in schizophrenics. *Psychological Reports,* 1962, **10,** 531–536.

SNYDER, S. Perceptual closure in acute paranoid schizophrenics. *Archives of General Psychiatry,* 1961, **5,** 406–410.

SNYDER, S., ROSENTHAL, D., & TAYLOR, I. A. Perceptual closure in schizophrenia. *Journal of Abnormal and Social Psychology,* 1961, **63,** 131–136.

SOLOMON, A. P., DARROW, C. W., & BLAUROCK, M. Blood pressure and palmer sweat (galvanic) responses of psychotic patients before and after insulin and metrazol therapy. *Psychosomatic Medicine,* 1939, **1,** 118–137.

SOMMER, R., DEWAR, R., & OSMOND, H. Is there a schizophrenic language? *Archives of General Psychiatry,* 1960, **3,** 665–673.

SOMMER, R., WITNEY, GWYNNETH, OSMOND, H. Teaching common associations to schizophrenics. *Journal of Abnormal and Social Psychology,* 1962, **65,** 58–61.

SPENCE, JANET T., & LAIR, C. V. Associative interference in the verbal learning performance of schizophrenics and normals. *Journal of Abnormal and Social Psychology,* 1964, **68,** 204–209.

SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *Journal of Experimental Psychology,* 1953, **45,** 265–277.

SUTTON, S., HAKEREM, G., ZUBIN, J., & PORTNOY, M. The effect of shift of sensory modality on serial reaction-time: A comparison of schizophrenics and normals. *American Journal of Psychology,* 1961, **74,** 224–232.

SYZ, H. C. Psychogalvanic studies in schizophrenia. *Archives of Neurology and Psychiatry,* 1926, **16,** 747–760.

SYZ, H. C., & KINDER, E. F. Electrical skin resistance in normal and in psychotic subjects. *Archives of Neurology and Psychiatry,* 1928, **19,** 1026–1035.

TAYLOR, JANET A., & SPENCE, K. W. Conditioning level in the behavior disorders. *Journal of Abnormal and Social Psychology,* 1954, **49,** 497–502.

TIZARD, J., & VENABLES, P. H. Reaction time responses by schizophrenics, mental defectives, and normal adults. *American Journal of Psychiatry,* 1956, **112,** 803–807.

TIZARD, J., & VENABLES, P. H. The influence of extraneous stimulation on the reaction time of schizophrenics. *British Journal of Psychology,* 1957, **48,** 299–305.

TOURNEY, G., FROHMAN, C. E., BECKETT, P. G. S., & GOTTLIEB, J. S. Biochemical mechanisms in schizophrenia. In R. Roessler and N. S. Greenfield (Eds.), *Physiological correlates of psychological disorder.* Madison: Univer. Wisconsin Press, 1962.

VENABLES, P. H. Factors in the motor behavior of functional psychotics. *Journal of Abnormal and Social Psychology,* 1959, **58,** 153–156.

VENABLES, P. H. The effect of auditory and visual stimulation on the skin potential response of schizophrenics. *Brain,* 1960, **83,** 77–92.

VENABLES, P. H. Changes due to noise in the threshold of fusion of paired light flashes in schizophrenics and normals. *British Journal of Social and Clinical Psychology,* 1963, **2,** 94–99. (a)

VENABLES, P. H. Selectivity of attention, withdrawal, and cortical activation. *Archives of General Psychiatry,* 1963, **9,** 74–78. (b)

VENABLES, P. H., & O'CONNOR, N. Reaction times to auditory and visual stimulation in schizophrenic and normal subjects. *Quarterly Journal of Experimental Psychology,* 1959, **11**, 175–179.

VENABLES, P. H., & TIZARD, J. The effects of stimulus light intensity on the reaction time of schizophrenics. *British Journal of Psychology,* 1956, **47**, 144–145. (a)

VENABLES, P. H., & TIZARD, J. Paradoxical effects in the reaction time of schizophrenics. *Journal of Abnormal and Social Psychology,* 1956, **53**, 220–224. (b)

VENABLES, P., & TIZARD, J. Performance of functional psychotics on a repetitive task. *Journal of Abnormal and Social Psychology,* 1956, **53**, 23–26. (c)

VENABLES, P. H., & TIZARD, J. The effect of auditory stimulus intensity on the reaction time of schizophrenics. *Journal of Mental Science,* 1958, **104**, 1160–1164.

VENABLES, P. H., & WING, J. K. Level of arousal and the subclassification of schizophrenia. *Archives of General Psychiatry,* 1962, **7**, 114–119.

WECKOWITZ, T. E. Size constancy in schizophrenic patients. *Journal of Mental Science,* 1957, **103**, 475–486.

WECKOWICZ, T. E. Autonomic activity as measured by the Mecholyl test and size constancy in schizophrenic patients. *Psychosomatic Medicine,* 1958, **20**, 66–71.

WECKOWICZ, T. E. Perception of hidden pictures by schizophrenic patients. *Archives of General Psychiatry,* 1960, **2**, 521–527.

WECKOWICZ, T. E. Shape constancy in schizophrenic patients. *Journal of Abnormal and Social Psy-Psychology,* 1964, **68**, 177–183.

WECKOWICZ, T. E., & BLEWITT, D. B. Size constancy and abstract thinking in schizophrenic patients. *Journal of Mental Science,* 1959, **105**, 909–939.

WECKOWICZ, T. E., & HALL, R. Distance constancy in schizophrenics and non-schizophrenic mental patients. *Journal of Clinical Psychology,* 1960, **16**, 272–276.

WECKOWICZ, T. E., SOMMER, R., & HALL, R. Distance constancy in schizophrenic patients. *Journal of Mental Science,* 1958, **104**, 1174–1182.

WENGER, M. A., JONES, N. F., & JONES, M. H. *Physiological psychology.* New York: Holt, 1956.

WENGER, M. A. *Evaluation of project 6, summary of proceedings: second cooperative psychological research conference.* Cincinnati: Veterans Administration Hospital, 1959.

WHATMORE, C. B., & ELLIS, R. M., JR. Some motor aspects of schizophrenia: An EMG study. *American Journal of Psychiatry,* 1958, **114**, 882–889.

WIENCKOWSKI, L. A. Stimulus factors influencing the disjunctive reaction time of schizophrenic and "normal" subjects. Unpublished doctoral dissertation, University of Buffalo, 1959.

WILDER, J. The law of initial values. *Psychosomatic Medicine,* 1950, **12**, 392–401.

WILLIAMS, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosomatic Medicine,* 1953, **15**, 456–462.

WISHNER, J. The concept of efficiency in psychological health and in psychopathology. *Psychological Review,* 1955, **62**, 69–80.

WISHNER, J. Efficiency: Concept and measurement. In, Proceedings of the XIVth International Congress of Applied Psychology, Vol. 2. *Personality research.* Copenhagen: Munksgaard, 1962. Pp. 161–187.

WYNNE, R. D. Can schizophrenics give the associations that "most people" do? Paper read at Eastern Psychological Association, New York, April 1963.

ZAHN, T. P., ROSENTHAL, D., & LAWLOR, W. G. GSR orienting reactions to visual and auditory stimuli in chronic schizophrenic and normal subjects. *Psychophysiological Newsletter,* 1963, **9**, 43–50.

ZAHN, T. P., ROSENTHAL, D., & SHAKOW, D. Reaction time in schizophrenic and normal subjects in relation to the sequence of series of regular preparatory intervals. *Journal of Abnormal and Social Psychology,* 1961, **63**, 161–168.

ZAHN, T. P., ROSENTHAL, D., & SHAKOW, D. Effects of irregular preparatory intervals on reaction time in schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **67**, 44–52.

ZAHN, T. P., SHAKOW, D., & ROSENTHAL, D. Reaction time in schizophrenics and normal subjects as a function of preparatory and intertrial intervals. *Journal of Nervous and Mental Diseases,* 1961, **133**, 283–287.

(Received March 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 107–110

EFFECTS OF PERCEPTUAL TRAINING ON UNMATCHED SAMPLES OF BRAIN-INJURED AND FAMILIAL RETARDED CHILDREN

DAVID ELKIND
University of Denver

RONALD R. KOEGLER, ELSIE GO
University of California, Los Angeles

AND WILLIAM VAN DOORNINCK
University of Denver

41 brain-injured (BI) and 39 familial retarded (FR) children matched for CA but not for MA were 1st tested on 1 set of ambiguous pictures, then trained on a 2nd set, and then retested on the 1st set both immediately and 1 month following training. Results showed that although the BI group had a significantly higher IQ than the FR group, they nevertheless took more trials to reach a lower level of posttraining performance. The difference between the groups reached statistical significance, however, only when the MA differential favoring the BI group was controlled by the analysis of covariance procedure.

The observation that brain-injured (BI) children experience greater difficulty on perceptual tasks than do familial retarded (FR) children of comparable CA and MA has been reported by many different investigators. It has been demonstrated, for example, that compared to the FR children, BI youngsters are less susceptible to the experience of apparent movement (Werner & Thuma, 1942), more subject to interference from background stimulation (Werner & Strauss, 1941), and less subject to the impression of figure-ground reversal (McMurray, 1954). Other results in a similar vein have recently been reviewed by Gallagher, 1957.

The major objection to these observations lies in the initial matching of the two groups with respect to MA and CA, i.e., with respect to IQ. Although IQ matching is a generally accepted procedure for investigations of this type, such matching is only appropriate when the IQ distributions of the parent populations have ranges of equivalent magnitude. This is patently not the case for the two groups in question, however, since the range of the FR group is curtailed by definition, whereas it has been estimated that 40% of the BI group have IQs of 100 or better. This means that it would not be possible to really match the IQ scores of the two populations as a whole since only their upper and lower tails overlap. Accordingly, it is not appropriate to match samples from these two populations since this matching pairs the brightest of the FR group with the dullest of the BI group. Results obtained from studies in which IQ matching has been employed are, therefore, not generalizable to BI and FR populations as a whole.

In order to say that the entire brain-injured population is perceptually handicapped in comparison to the entire retarded population one would have to compare the perceptual abilities of representative (unmatched) samples. With unmatched samples the experimental question would be to determine whether the IQ differential favoring the BI population is sufficient to compensate for the perceptual differential favoring the FR population. The purpose of the present study was to obtain a preliminary answer to this question by comparing unmatched samples of BI and FR children for their ability to profit from training in the perception of ambiguous figures, both with and without statistical control for the MA differential between the groups.

METHOD

Subjects. Forty-one children from the patient population of the Kennedy Foundation Clinic in Los Angeles served as the brain-injured sample. Criteria for selection as a BI child were a diagnosis

TABLE 1

MEANS, STANDARD DEVIATIONS, AND *t*'S FOR AGE AND IQ FOR BRAIN-INJURED (BI) AND FAMILIAL RETARDED (FR) CHILDREN

Group	*N*	Age (in months)		IQ[a]	
		M	*SD*	*M*	*SD*
BI	41	93.37	17.97	87.41	14.26
FR	39	101.49	25.03	76.10	8.86
t		1.64 *ns*		4.22**	

[a] Stanford Binet.
* $p \leq .01$.

of cerebal palsy (CP) and a noticeable motor impairment.[1] Thirty-nine children attending classes for the educable retarded in the Los Angeles public schools served as the familial retarded sample. Criteria for selection as an FR child were the absence from the record of any diagnosis of brain damage and the absence of any case history material suggestive of it. The children were roughly matched for CA but not for MA. Table 1 gives the means and standard deviations for age together with the means and standard deviations for IQ (Stanford Binet). As expected, Table 1 indicates that there was a significant IQ difference between the unmatched samples which favored the BI group.

Materials. The materials consisted of two sets of ambiguous figures printed in black and white and mounted on 8 × 11½ inch tagboard sheets. For one set of figures cardboard shields were available which were so cut that when they were placed over the ambiguous figure the hidden (reversed) figure became immediately apparent (Figure 1). The average correlation between untutored performances on the two sets of drawings, scored as described below, was .50 ($p < .05$), (Elkind, in press).

Procedure. The procedure was the same as that used in a previous study with normal children (Elkind, Koegler, & Go, 1962). Each child was pretested on one set of cards (Set A) immediately prior to training. The testing consisted of showing the child the cards one at a time and asking him, "What do you see? What does it look like?" Immediately after this pretest the child was trained on the second set of cards (Set B). Training involved providing the child with successively more direct and revealing clues to the perception of the hidden figures. The first clue was the statement, "Some children see more than one thing. Do you see anything else besides a ——?" (whatever the child had seen). If this clue did not prompt recognition of the hidden figure, a second verbal clue was given. "Sometimes children see a —— (whatever the child had *not* seen) in the cards. Do you see a ——?" If the child replied that he did see the new figure, he was asked to point out its parts to insure that he was not merely responding to suggestion. Those children who still could not see the figure were then given a third, *perceptual* clue. This time the cardboard shield was superimposed upon the picture and the previous question was repeated. The youngster was again asked to point out the parts to insure that he actually saw the hidden figure. Each child was trained to the point that he could indicate the parts of all the hidden figures on Set B. Immediately after the training session the child was again tested on Set A and then retested on Set A 1 month later. The child's score for each testing was the number of perceived figures of the 21 possible figures contained in the drawings of Set A. Each child was also given a learning score which was the number of clues he required to see all of the hidden figures in Set B.

The results of the experiment were first assessed by a two-way analysis of variance (for the three testings on Set A and for the two diagnostic groups) and then by an analysis of covariance with MA as the control variable. An analysis of covariance was also used to determine the significance of the learning score differences between the BI and FR children.

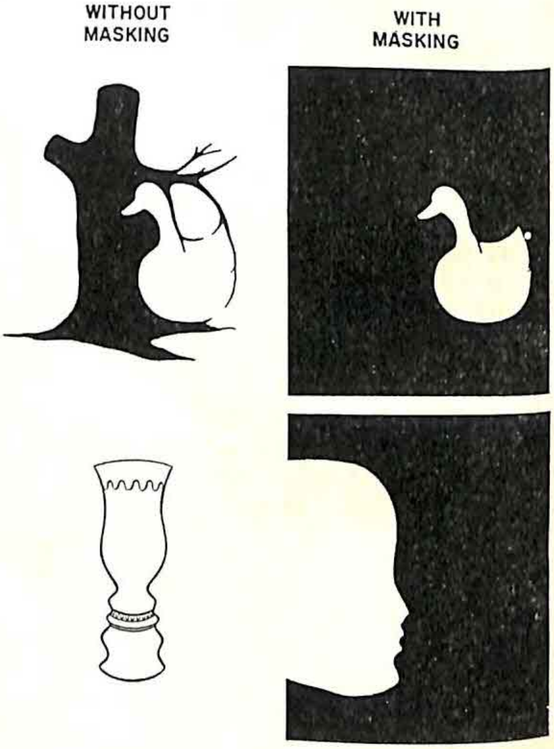


FIG. 1. Illustration of the training procedure showing two of the figures with and without masking shields.

---

[1] This is a stringent criterion of brain injury but diagnosis of BI is often equivocal and we wanted to insure that our BI group was in fact brain injured. This criterion does, however, limit the representativeness of the BI sample and hence the generality of our results.

TABLE 2

CORRECTED (C) AND UNCORRECTED (UC) MEANS AND t'S FOR BRAIN-INJURED AND RETARDED CHILDREN FOR THREE TESTINGS ON SET A

Group	N	Pretest		Immediate Posttest		1 month Posttest	
		UC	C	UC	C	UC	C
BI	41	6.53	6.47	9.88	9.52	9.15	8.77
FR	39	6.73	6.81	10.73	11.10	10.47	10.86
t			.80 *ns*		3.16*		4.19*

* $p \leq .01$.

RESULTS

Analysis of variance for testings and groups. The obtained F for the three testing on Set A was 10.58 ($p < .01$). Comparison of individual mean scores (t tests) revealed that both the immediate and 1 month posttraining scores were significantly higher than the initial pretraining values. On the other hand, although the FR group scored higher on both the immediate and 1 month posttraining measures, the differences between the two groups did not reach statistical significance and the F for groups was only 1.25 ($p > .05$). The interaction of testing and groups was also nonsignificant.

Analysis of covariance for testings, groups, and MA. As revealed by the analysis of covariance, the F for testings was again statistically significant and mean score differences indicated significant improvement in ambiguous picture perception as a result of training. The control of MA, however, magnified the amount of improvement made by the FR group while it did just the reverse for the improvement made by the BI youngsters. Furthermore, in contrast to the nonsignificant F for groups obtained by the analysis of variance above, the covariance F for groups was 13.95 and was significant at the .01 level.

The significant F for groups obtained by the covariance procedure was attributable to the removal of the variance due to MA from the ambiguous picture scores. That is to say, the perceptual handicap of the BI group in comparison to the FR group was to a large extent compensated by the BI group's significantly higher mean IQ. When the effects of this compensating factor were removed, the difference favoring the FR children became statistically significant. Table 2 presents the mean scores for the two groups on all three testings both with and without correction for the correlation (average for three testings +.45) between ambiguous picture scores and MA. In Table 2 tests for differences between the means are presented only for the corrected means because the F for the uncorrected ones did not reach statistical significance.

Analysis of covariance for learning trials. For the BI group the mean number of trials required to perceive all the hidden figures on Set B was 20.79, whereas for the FR group the mean was 19.33. When the effects of MA were removed by the covariance procedure, the adjusted BI group mean was 21.4 while the adjusted FR group mean was 18.72. The difference between the uncorrected means for learning trials was not significant, but the difference between the corrected values was significant at the .05 level. The correlation between the number of learning trials required and MA was −.51.

DISCUSSION

Results of the present study have shown that unmatched samples of BI and the FR children made significant improvement in ambiguous picture perception as a result of perceptual training and that the effects were apparent both immediately and 1 month after training. In addition, it was found that despite its significantly higher mean MA, the BI group required a few more training trials and reached a slightly lower level of ambiguous picture perception than did the FR subjects. When statistical control for the MA advantage of the BI group was introduced, the differences between the two groups were substantially widened and reached statistical significance. An incidental finding was that for the ambiguous pictures used in the present study learning and posttraining scores discriminated between the two groups whereas the initial performance (pretest) scores did not.

These results suggest that although the perceptual handicap of BI children can be demonstrated in comparison to FR children, it is not of sufficient magnitude to reach statistical significance without the elimination of the MA difference between the groups. Such elimination is, however, really not defensible since, as we have argued, the two groups do in fact differ with respect to their mean MAs so that control of the MA differential creates an equivalence that does not exist in fact. It is, therefore, not correct to say that the BI child is significantly poorer in perceptual ability than the familial retarded child of comparable MA because in a majority of instances the MAs would not be comparable. All we are really justified in concluding, at least for CP injured youngsters, is that: the BI child requires *a higher* level of mental ability than the FR child to reach *the same* level of perceptual achievement.

CONCLUSION

It was concluded that: (*a*) the widely-accepted practice of matching BI and FR groups for MA and CA introduces a systematic bias of sampling which makes the perceptual handicap of the BI group, in comparison to the FR group, appear much larger than it is in reality; and (*b*) the real difference between the BI and FR children consists in the fact that the BI child requires *a higher level* of mental ability than the FR child to reach *the same level* of perceptual performance.

REFERENCES

ELKIND, D. Ambiguous pictures for the study of perceptual development and learning. *Child Development*, in press.

ELKIND, D., KOEGLER, R. R., & GO, ELSIE. Effects of perceptual training at three age levels. *Science*, 1962, **137**, 755–756.

GALLAGHER, J. J. A comparison of brain-injured and non-brain-injured mentally retarded children on several psychological variables. *Monographs of the Society for Research in Child Development*, 1957, **65**, 1–79.

MCMURRAY, J. S. Rigidity in conceptual thinking in exogenous and endogenous mentally retarded children. *Journal of Consulting Psychology*, 1954, **18**, 366–370.

WERNER, H., & STRAUSS, A. A. Pathology of figure background relation in the child. *Journal of Abnormal and Social Psychology*, 1941, **36**, 58–67.

WERNER, H., & THUMA, B. D. A deficiency in the perception of apparent motion in children with brain injury. *American Journal of Psychology*, 1942, **55**, 58-67.

(Received February 12, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 111-113

ACQUIESCENCE RESPONSE SET IN APHASICS[1]

JAMES J. JENKINS
University of Minnesota
AND HILDRED SCHUELL
University of Minnesota Medical School, Minneapolis Veterans Administration Hospital

2 tests of sentence comprehension and 2 tests requiring detection of word errors in sentences were administered to normal and aphasic patients. Each sentence could be responded to with "yes" or "no" or "right" or "wrong." Sentence content involved simple facts known to most adults. Little evidence of acquiescence response set was found for normals, but aphasics showed significant and marked acquiescence response bias on all tests.

A commonplace observation with regard to deaf and brain-damaged populations is that many patients adopt an "agreement set" in order to minimize the extent to which they display their comprehension difficulties and, perhaps, to make the social situation less awkward for others even though they do not understand it themselves. To the best of the writers' knowledge, little objective work has been performed to furnish support for this observation. Recently the literature has shown concern with response sets in other situations and in other populations (see Milholland, 1964, for a review), and a variety of such sets have been examined, with "acquiescence set" being the most relevant to the present study.

The purpose of this paper is to report on the acquiescence set observed in a group of aphasic patients. It is the more remarkable in that the behavior was manifested in a test situation rather than in a simple social situation and that the items were factual items, the truth or falsity of which is well-known in the adult population from which the aphasics were drawn.

METHOD

Four tests of simple language skills were devised. Each test had 15 items. Two tests designed to measure sentence comprehension consisted of a series of sentences which could be answered with a simple indication of "yes" or "no," e.g. "Is an orange a fruit?" "Are towns bigger than cities?" "Does the sun rise in the west?" "Do eggs come from chickens?" One of these tests was presented auditorially and the other in written form. The auditory form was answered orally or by gesture; the written form by checking "yes" or "no." In each test positive answers were correct for eight sentences and negative answers were correct for seven. The other two tests involved the detection of errors which consisted of the substitution of an irrelevant word for a relevant word in a sentence. In the auditory form a statement was read aloud to the patient and he was asked if it was correct, e.g. "You eat with a knife, fork, and comb. Is that right?" "The American flag is red, white, and blue. Is that right?" In the visual form the patient read the item himself and checked either "Right" or "Wrong," e.g., The population of the United States is increasing .... Right ( ) Wrong ( ). You wash your hands with soap and water .... Right ( ) Wrong ( ). In each of these tests 7 items were right and 8 items were wrong.

Subjects. The four tests were individually administered to aphasic patients and controls. The control sample of 30 normal, non-brain-damaged adults was chosen from patients on the medical wards of the Minneapolis Veterans Administration Hospital. The subjects were selected to match the age distribution of aphasic patients. The aphasic subjects were drawn from patients referred to the Aphasic Division of the Neurology Service of the Minneapolis Veterans Administration Hospital for testing. These were consecutive admissions except that patients without aphasia, patients who could not respond to test items and psychotic patients were excluded. The age range was from 21 to 70. Etiologies were diverse with cerebrovascular accidents as the single most frequent source. The number of patients given each test varied from 51 to 60 because these experimental tests were put into use at different dates, and tests were administered subject to other demands on the patients' time.

RESULTS

Table 1 presents the number of errors per subject per item in the normal and the aphasic groups for each of the sets of items on each test. It is apparent that the aphasics per-

---

[1] This research was supported in part by Grant B-1948 from the Institute of Neurological Diseases and Blindness, United States Public Health Service.

formed much more poorly than normals on all of the tests, as would be expected. On the first three tests there is no item on which the controls make as high a proportion of errors as the aphasics. Only in the last test in which the errors are very much of the form of proofreader errors which normals have been trained to overlook (e.g., *soup* for *soap, stones* for *stores*) are there any items in which control errors reach or exceed aphasic errors.

With respect to response bias in favor of the acquiescence set on the first three tests the normals show virtually no tendency to make more errors on items on which the correct answer is "no" or "wrong." In the last test where the proofreader errors show a high rate of nondetection there is an appreciable bias present.

The aphasics, however, show a marked and consistent tendency on all of the tests to make errors by saying "yes" or "right" when the correct answer is "no" or "wrong" rather than making errors in the reverse direction. The strength of this tendency runs from almost 4.4 to 1 on the first test (Comprehension of Auditory Sentences) to about 1.8 to 1 on the second test (Comprehension of Written Sentences). Means and standard deviations for aphasic errors on both kinds of sentences are given in Table 2 for each test.

A test of the significance of the differences in mean errors on the two sentence types was performed for the aphasic data for each test separately. Because variances differed for the two sentence types on some of the tests, a conservative t test was employed as suggested by Dixon and Massey (1957, pp. 123–124). Results of t tests are given in Table 2. The differences between errors occurring on the two sentence types are significant beyond the .05 level for each test.

TABLE 1

AVERAGE ERROR PER SENTENCE PER SUBJECT ON FOUR LANGUAGE TESTS

Subjects	Test	N	Positive answer correct	Negative answer correct
Normal	1	30	0.000	0.024
	2	30	0.013	0.033
	3	30	0.010	0.004
	4	30	0.000	0.300
Aphasic	1	51	0.069	0.303
	2	54	0.111	0.199
	3	60	0.074	0.140
	4	52	0.124	0.313

TABLE 2

MEANS AND STANDARD DEVIATIONS OF APHASIC ERRORS BY SENTENCE TYPES ON THE FOUR TESTS

Test	N[a]	Positive answer correct		Negative answer correct		t	p
		M	SD	M	SD		
1	51	3.50	1.14	15.43	2.50	7.23	<.01
2	54	6.00	3.93	10.14	5.47	3.80	<.01
3	60	4.57	2.91	8.38	3.42	2.30	<.05
4	52	6.43	2.50	16.25	7.94	3.35	<.01

[a] Patients.

DISCUSSION

The results clearly indicate that the aphasic patients show an appreciable bias toward responding in the positive direction (agreeing with a sentence or saying that it is correct) as opposed to distributing their errors randomly. The same tendency is scarcely detectable in the control group of normals except in the case of the recognition of visual errors which closely resemble the correct word which would make the sentence true. In this latter case normals "misread" the word in such a fashion that the sentence can be responded to as true. Such misreading is, of course, well known and probably constitutes a useful habit in almost all reading tasks (except the test task and actual proofreading, of course). While such errors may appear to be qualitatively different from the other kinds of errors studied here, in principle they may be viewed as similar. The proofreading errors occur because other stimulus elements in the situation (here the context of the sentence) make the replaced word a highly probable one and the stimuli from the replacing word are insufficiently different from those of the replaced word to defeat the expectation. It is noteworthy that normals do not misread a correct sentence as incorrect. All errors are in a conforming, probabilistic direction.

For the aphasics all of the tests may be seen to involve something of the same nature. The perception of individual word cues and order cues and the discrimination of a word from its close associate or semantic similar (e.g., *east* from *west*, *all* from *some*) are set against high probability responses of agreement and the general high probability of the word clusters which the aphasic does perceive (e.g., towns, cities, are, bigger, than). Hence his most likely response is agreement.

A lesson for test builders is also implied here. If one wishes to discriminate between the aphasic and normal samples employed here, the items to which negative responses are correct are more powerful than are the items to which a positive response is correct (with the exception of the test for the recognition of word errors in writing).

In conclusion it appears that the general observation of agreement tendencies in a brain-damaged population is well supported by the data. Aphasics adopt a strategy of agreement when dealing with materials which they do not completely understand.

REFERENCES

DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1957.

MILHOLLAND, J. E. Theory and techniques of assessment. *Annual Review of Psychology*, 1964, **15**, 311–346.

(Received February 19, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 114–118

RESPONSE TO SEXUALLY AROUSING STIMULI AS A FUNCTION OF REPRESSING AND SENSITIZING DEFENSES[1]

DONN BYRNE AND JOHN SHEFFIELD
University of Texas

It was hypothesized that individuals who utilize sensitizing defenses respond to sexually arousing stimuli with greater verbalized anxiety than do individuals who utilize repressing defense mechanisms. From a large subject pool, 44 repressers and 44 sensitizers were selected on the basis of scores on the Repression-Sensitization Scale. Half of the subjects read a series of vividly descriptive sexual passages, and half read a series of neutral passages. Self-ratings were obtained immediately afterward. Sex arousal was significantly greater for the experimental groups. As predicted, sensitizers were more anxious than repressers in the sex arousal condition, but the 2 groups did not differ in the neutral situation. Sex arousal was associated with feelings of anxiety, entertainment, and lack of boredom among sensitizers and with disgust and anger among repressers.

It is apparent with respect to a great many specific motivational states that anxiety is a concomitant of drive arousal. A sufficiently prolonged state of arousal of most primary drives represents a sufficient degree of threat to arouse anxiety. Human beings, however, have a curious propensity for associating anxiety with motivational states which do not in themselves lead to bodily harm. It has been found, for example, that the experimental arousal of affiliation need elicits anxiety (Byrne, 1961a). To an even greater extent, one would expect sexual stimuli to be anxiety arousing in our culture. In a great many otherwise diverse investigations, threat has simply been defined in terms of exposure to sexual words, pictures, or recordings (e.g., Aronson & Mills, 1959; Lazarus, Eriksen, & Fonda, 1951; McGinnies, 1949; Shannon, 1962).

In a few instances, experimenters have manipulated sex arousal experimentally. Clark (1952) exposed male subjects to sexually relevant stimuli (photographic slides of nude females in one group and an attractive female experimenter in another) and found significantly *less* manifest sex imagery in their TAT protocols than in those of control groups. An interpretation of this finding in terms of inhibitory anxiety is supported by the positive association between the expression of sex imagery and guilt over sexual activity (Clark, 1952) and by the increase in symbolic sex imagery in the arousal groups (Clark & Sansibar, 1955). Further, the ingestion of alcohol by subjects viewing the nude slides resulted in significantly more manifest and less symbolic sex imagery than in the control group, presumably because of a reduction in anxiety (Clark, 1952). Mussen and Scodel (1955) attempted to vary inhibitory cues by the use of a stern father figure versus a young, permissive experimenter to administer photographic slides of nude females. As hypothesized, significantly less sexual imagery was present in the TAT stories of the subjects in the former group than in the stories of the latter; the authors suggest that the father figure served to increase the amount of sex-associated anxiety. Martin (1964) employed the amount of time spent in sorting pictures of nude females as the dependent variable and experimentally created inhibitory and permissive conditions. Sorting time was significantly shorter in the former than in the latter condition. In the second experiment, he found greater increases in palmar skin conductance when the nude pictures were presented under inhibitory than under permissive conditions. Thus, the notion that inhibitory cues lead to anxiety and

---

[1] The authors wish to thank James Barry who served as a research assistant for a portion of this investigation.

hence to differential subsequent behavior was supported.

It appears that sexually arousing stimuli frequently evoke anxiety, the amount of anxiety may be increased or decreased by the presence of other variables, and that the subsequent expression of sexually relevant responses is some function of the difference between the sexual motive and the anxiety drive.

In the research which has been conducted in this area, the possibility of individual differences in the amount of anxiety aroused by sexual stimuli or in the mechanisms utilized to deal with anxiety has been explored very little. Using a sex-guilt questionnaire, Leiman and Epstein (1961) found that individuals with high sex guilt gave fewer sex responses to sexually relevant TAT cards than did those with low sex guilt, though the two groups did not differ in responding to nonsexual cards. A considerable amount of research has been conducted on the repression-sensitization dimension which indicates its probable relevance to such behavior. This personality dimension is defined in terms of characteristic responses to anxiety-evoking stimuli. At the repressing extreme are behaviors which involve avoidance of anxiety-arousing stimuli and their consequents; at the sensitizing extreme are behaviors which involve an attempt to approach or control the stimulus and its consequents.

Much of the recent work in this area has utilized the Repression-Sensitization (R-S) Scale which is made up of MMPI items (Altrocchi, Parsons, & Dickoff, 1960; Byrne, 1961b; Byrne, 1964; Byrne, Barry, & Nelson, 1963). The construct validity of the R-S Scale as a measure of defense mechanisms is supported by a diverse array of findings. For example, expert judges, asked to respond to the test as a represser would, corresponded 90% with the scoring key (Byrne et al., 1963). Davison (1963) factor analyzed responses to a number of measures of coping mechanisms; the first factor was essentially defined by the R-S Scale, and the factor included intellectualization as judged in an interview situation. Utilizing Freudian theory, O'Connell and Peterson (1964) correctly hypothesized that appreciation of humor would be greatest for normals, least for repressers, and that sensitizers would fall between the other two groups.

Of more direct relevance to the present investigation, scores on the R-S Scale have also been found by a number of investigators to be predictive of response to threat. Compared to sensitizers, repressers are reported to have greater difficulty in recalling nonsense syllables associated with poor performance on an Intelligence and Personality Test (Gossett, 1964) and to have higher perceptual thresholds for words associated with failure (Tempone, 1962). In describing their own emotional response to threat, repressers tend to deny and sensitizers freely to admit feelings of anxiety (Davison, 1963; Lazarus & Alfert, 1964; Pomeranz, 1963). Whether the personality dimension measured by the R-S Scale is best conceptualized as a measure of defense mechanisms is, as with any new test, still an open question. It has been suggested elsewhere (Byrne, 1964) that "at the present time the label 'repression-sensitization' does not appear inconsistent with what is known about the R-S Scale."

In summary, then, sexual stimuli are assumed to constitute a threat, and individuals differing in repression-sensitization should respond differently to that threat. Specifically, it is hypothesized that sensitizers respond to sexually arousing stimuli with greater verbalized anxiety than do repressers.

PROCEDURE

The R-S Scale was administered to a group of approximately 150 undergraduate males enrolled in introductory psychology at the University of Texas. From this subject pool, 44 repressers ($M = 41.55$, Range = 31–48) and 44 sensitizers ($M = 77.60$, Range = 67–105) were selected from the upper and lower thirds of the distribution.[2]

Several weeks later, the subjects were assigned to one of two experimental conditions following a procedure for sex arousal which has been used previously (Byrne, 1957). All subjects were given a booklet entitled "Current Fiction" with the instructions, "Please read each of the following passages carefully. When you have finished all eleven of the selections, you will be asked several questions about what you have read." Each selection consisted of several paragraphs from *The Naked and the Dead*, *Ulysses*, *Peyton Place*, *The Revolt of Mamie Stover*,

---

[2] Scores are based on the original R-S Scale scoring system (Byrne, 1961b).

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SELF-RATINGS OF SEX AROUSAL

	Conditions			
	Sex arousal		Neutral	
	M	*SD*	*M*	*SD*
Repressers	3.05	.82	1.14	.34
Sensitizers	3.32	.82	1.05	.21

TABLE 3

MEANS AND STANDARD DEVIATIONS OF SELF-RATINGS OF ANXIETY

	Conditions			
	Sex arousal		Neutral	
	M	*SD*	*M*	*SD*
Repressers	1.91	.67	1.59	.72
Sensitizers	2.91	1.31	1.55	.84

Return to Peyton Place, God's Little Acre, Up in Michigan, The Ninth Wave, The Secret of Sylvia, Lady Chatterley's Lover, and *Aphrodite.* For half of the subjects, these passages consisted of vividly descriptive sexual content, while for the other half of the subjects, the passages were neutral scenes from the same books.

Immediately after reading the literary selections, the subjects were asked to respond to six five-point rating scales dealing with their feelings while reading the literary selections. They were asked to indicate the extent to which they were sexually aroused, disgusted, entertained, anxious, bored, and angry.

RESULTS

The efficacy of the arousal procedure was determined by comparing responses to the Sex Arousal scale in the two conditions for repressers and sensitizers. The scale ranged from "strongly aroused sexually" (five points) to "not at all aroused sexually" (one point). The means and standard deviations of these self-rating scores are presented in Table 1. As shown in Table 2, the only significant effect indicated by analysis of variance[3] was for the experimental conditions. Sex arousal was significantly greater following the sexual passages than the neutral passages.

In order to test the hypothesis, the self-ratings of anxiety were similarly analyzed. This scale ranged from "extremely anxious" (five points) to "not at all anxious" (one point), and the means and standard deviations are shown in Table 3. In the analysis of variance, both main effects and the interaction were found to be statistically significant, as may be seen in Table 4. Essentially, the sensitizers indicate more anxiety in the sex arousal condition than do the repressers whereas the two groups do not differ in the neutral condition. The hypothesis was confirmed.

As an additional way of exploring the effects of sex arousal on the two defense groups, the correlates of the sex arousal ratings were determined separately for the 44 repressers and 44 sensitizers as shown in Table 5. In examining the significant relationships, it is clear that repressers and sensitizers exhibit distinctly different patterns of response to sex arousal. For sensitizers, sexual arousal is associated with an increase in anxiety and an indication of feeling more entertained and

TABLE 2

ANALYSIS OF VARIANCE OF SELF-RATINGS OF SEX AROUSAL

Source	*df*	*MS*	*F*
Conditions	1	96.18	240.56***
Repression-Sensitization	1	.18	.45
Interaction	1	.73	1.83
Within	84	.40	

$^{***}p < .001$.

TABLE 4

ANALYSIS OF VARIANCE OF SELF-RATINGS OF ANXIETY

Source	*df*	*MS*	*F*
Conditions	1	15.56	17.48**
Repression-Sensitization	1	5.01	5.63*
Interaction	1	6.01	6.75*
Within	84	.89	

$^{*}p < .05$.
$^{**}p < .01$.

---

[3] In each analysis of variance, Hartley's maximum *F* ratio test for heterogeneity of variance yielded a significant F max. Since it has been found, however, that departure from homogeneity has little influence upon the power function (Young & Veldman, 1963), the raw data were not transformed.

TABLE 5

CORRELATES OF SELF-RATINGS OF SEX AROUSAL

	Repressers	Sensitizers
	r	r
Disgusted	.34*	.06
Entertained	.02	.33*
Anxious	.25	.67**
Bored	−.15	−.57**
Angry	.36*	.00

* $p < .05$.
** $p < .01$.

less bored. Repressers, on the other hand, indicate that they feel more disgusted and angry as sexual arousal increases.

DISCUSSION

Perhaps the most important aspect of the reported findings lies in the delineation of the different responses associated with sex arousal for subjects differing in the characteristic way in which they respond to threat. Male college students, after being exposed to verbal material having sexual content, indicate feelings of sex arousal as a consequence. When these individuals are divided into groups based on their scores on a measure of defenses, the concomitants of arousal are quite different for the two groups. Why?

Obviously, a number of interpretations are possible. For example, it may be that sensitizers have greater conflicts concerning sexual behavior as shown by the presence of both positive and negative affect in response to sex cues. Repressers, conversely, could be those individuals who are relatively free of sexual conflict; exposure to sexual stimuli simply leads to sex arousal without any associated anxiety. There is a considerable amount of evidence which suggests that individuals identified as sensitizers by the R-S Scale are less well adjusted than individuals identified as repressers. It has been found that, compared with repressers, sensitizers give more deviant responses on Gough's Adjective Check List (Byrne, 1961b; Lucky & Grigg, 1964), have a larger self-ideal discrepancy (Byrne, 1961b; Byrne et al., 1963), are more anxious (Joy, 1963), and yield lower (more maladjusted) profiles on the CPI (Byrne, Golightly, & Sheffield, in press). Thus, the present findings may simply indicate that more anxious, maladjusted individuals respond differently to sexual threat than do well-adjusted individuals. Since the evidence relating to adjustment rests on verbal responses to paper and pencil tests and since repressers by definition respond to such tests by indicating a feeling of well being, an interpretation of the R-S Scale as a measure of adjustment is by no means clear-cut.

A different, and to the authors more likely, interpretation of the findings assumes that sexual stimuli are equally threatening to individuals at the two ends of the defense continuum. Sensitizers meet this threat characteristically by verbalizing their feelings of anxiety. In effect, they indicate awareness of their conflictful state in that they indicate the simultaneous presence of feelings of sexual excitement, anxiety, and pleasure. For the unmarried college student in our culture, this would appear to be a veridical characterization.

The problem, though, is to secure evidence that the represser subjects were also threatened but were unwilling or unable to verbalize feelings of anxiety. The only pertinent findings in the present data are the positive correlations between sex arousal, anger, and disgust for the repressers. It could be argued that these indications of hostile affect represent a disguised and indirect indication that sexual excitement is disturbing to them. A sequence of the following sort would be postulated:

Sexual Stimuli → Sex Arousal → Anxiety →
Denial of Anxiety → Frustration
→ Aggressive Responses

In other investigations, there is evidence supportive of the proposition that the denial of anxiety by repressors in threatening situations is accompanied by indirect indicators of disturbance. J. F. Lomont (cited by Byrne, 1964) administered the R-S Scale, the IPAT Self-Analysis Form, and a 66-item word association test to 35 hospitalized patients. He found that sensitizers tended to respond with greater verbalized anxiety as measured by the IPAT ($r = .76$), but that repressers tended to indicate greater disturbance (long reaction time, blocking, rehospitalization being 59.6 and 79.5 months for the remitted and nonremitted groups, respectively.[3]

Within the remitted and nonremitted groups, half were assigned to the experimental condition and half to control. The various subgroups were matched for age and education, the means for the total group being 37.5 years and 11.8 years for age and education, respectively. Eleven additional subjects (five remitted, six nonremitted) were seen but testing was not completed either because of the subject's refusal to continue or failure to achieve a criterion of 3 correct anticipations (out of a possible 60) on a practice list.

Procedure. The materials and experimental procedure were the same as those used in the previous study by the writers (Spence & Lair, 1964). Briefly, each subject was given two paired-associate lists. The first (practice) list consisted of six pairs of familiar words with minimal similarity and associative connection among them. Two second lists were prepared. The experimental list was formed by selecting eight of the stimulus words used by Russell and Jenkins (1954) in their word association study and, as responses, the most dominant association to each of these stimuli contained in their norms. Each response, however, was *not* paired with the stimulus to which it was an association but with another stimulus on the list. The same responses appeared on the control list. The stimuli were words beginning with the same letter and of the same length and familiarity level as the parallel stimulus in each pair on the experimental list. Associative connections between and within the pairs of this list were at a minimum.

After being shown a two-pair demonstration list, the subject was given 10 anticipation trials on List 1 followed by 15 anticipation trials on either the control or experimental form of List 2 or, in each instance, the number of trials required to reach two successive perfect recitations, whichever occurred first. The lists were presented to the subject manually at a 5:5 second rate by means of 3 × 5 index cards, each stimulus and stimulus-response pair appearing on a separate card.

RESULTS

A number of subjects, particularly those in the control condition, reached the learning criterion of two successive perfect recitations on List 2 well before the fifteenth anticipation trial. It was therefore decided to analyze the data from only the first five anticipation trial. It was therefore decided to analyze the the total number of correct responses given over these trials on List 1 revealed that the performance of the remitted subjects was significantly superior to that of the nonremitted ($F = 4.01$, $p < .05$). To take account of this difference in initial performance level, the List 2 data were analyzed by means of gain scores, computed for each subject by subtracting from the number of correct responses made on Trials 1–5 of List 2 the number made on the same trials of List 1.

TABLE 1

MEAN GAIN SCORES (LIST 2–1) FOR THE FOUR SUBGROUPS

	Remitted		Nonremitted	
	M	*SD*	*M*	*SD*
Experimental	11.30	5.12	6.65	10.51
Control	16.50	5.98	16.90	7.25

As may be seen from Table 1, the means of both the experimental groups were markedly lower than those of their controls. A 2 × 2 analysis of variance showed that the difference was highly significant ($p < .001$), thus demonstrating that the presence of the cross-associates on the experimental list retarded performance. It may also be observed that while the gain scores of the two control groups were approximately equal, the mean of the nonremitted experimental subjects was less than that for the remitted. This interaction approached statistical significance ($F = 3.35$; $.05 < p < .10$). The gain scores of the remitted and nonremitted groups were also compared with those of the normal subjects tested in the writers' previous study (Spence & Lair, 1964). In an analysis comparing the nonremitted groups with the normals, the interaction between group and list was found to approach statistical significance ($F = 3.02$, $p < .10$) but in the analysis involving the remitted groups it did not ($F < 1.00$).

Turning to the overt errors made by the two groups of the present study, inspection of the percentages of incorrect responses represented by overt intrusions on both List 1 and List 2 indicated that there were no significant differences between the remitted and nonremitted subjects. The percentages of both groups, however, were higher than those previously found with the normal subjects and approximately the same as had been found with the short-term schizophrenics (Spence & Lair, 1964).

---

[3] As indicated by a t test, this difference was statistically significant ($p < .05$). Inspection of the data showed, however, that length of hospitalization per se was unrelated to any of the experimental measures.

Although not differing in proportion of overt errors, the remitted and nonremitted groups did differ in their tendency to give extralist responses. On List 1, 27 of the 40 nonremitted subjects gave 1 or more of these responses, in contrast to 17 of the remitted ($x^2 = 5.05$, $p < .05$). The nonremitted subjects continued to give more extralist responses to both forms of List 2. Particularly marked was the occurrence of extraexperimental responses (those that had not been stimulus or response terms on either List 1 or 2), 19 of the nonremitted giving responses of this type and only 8 of the remitted ($x^2 = 4.02$, $p < .05$).

Several other observations concerning these errors are of interest. In both remitted and nonremitted groups, few of the extraexperimental responses were contained among the Russell-Jenkins (1954) list of associations to the stimulus words. Thus if these responses represented intrusions of preestablished associations, they were idiosyncratic ones. Secondly, in every subgroup the mean number of correct responses for subjects who gave extraexperimental errors was less than for subjects who gave none. Due in part to small *N*s none of these differences was statistically significant although significance was obtained for the nonremitted subjects on List 1 when the experimental and control groups were combined ($t = 2.37$, $p < .05$). Finally, the mean gain score of experimental subjects who gave extraexperimental responses was found not to differ significantly ($t < 1.00$) from the mean of subjects who did not, thus giving no indication that the former were more susceptible to associative interference. Similar results have been obtained with college students (Spence, 1963), subjects giving extralist responses exhibiting a generally poorer performance but no greater vulnerability to intralist associative interference than subjects who gave none.

DISCUSSION

The interaction that was found between the current mental status of the patients and type of list in determining performance provides evidence that at least sicker, relatively nonremitted schizophrenics tend to exhibit exaggerated associative interference of a type predicted by Mednick's (1958) theory. The results do not, however, provide any direct confirmation of his hypotheses concerning the factors leading to the *appearance* of such associative disturbances. According to Mednick's views, various anomalies in the intellectual functioning of schizophrenics are originally the direct result of a high level of anxiety or drive but with time come to be learned reactions so that they continue to be observed even in chronic patients with flat affect. Tests of these latter hypotheses would seem to demand the conduct of experiments showing directly that anxiety or drive level is related to degree of associative interference in the predicted manner, of investigations in which the same schizophrenic subjects are tested at various stages of the disorder, etc. The positive results of the present study do, however, suggest the potential usefulness of investigating these aspects of Mednick's theorizing.

Consideration of the data obtained from the error analyses raises the possibility, however, that associative interference of the type suggested by Mednick and manipulated here may not be the most appropriate measure to use in future investigations. A tendency to give extralist responses and a tendency to exhibit heightened associative interference can each be considered as an example of associative disturbance. The failure in the present investigation to find that one covaries with the other suggests that they may not be due to a common set of individual difference factors. In this and the writers' previous study, the error measure consistently differentiated between groups (schizophrenics and normals, remitted and nonremitted schizophrenics) while evidence for abnormal intralist associative interference was found only in nonremitted patients and then only at a borderline level of statistical significance. It may be that the factors responsible for the associative anomalies reflected in the error measure are more closely related to the schizophrenic disorder than those responsible for the appearance of heightened intralist associative interference.

REFERENCES

CARSON, R. C. Intralist similarity and verbal rote learning performance of schizophrenic and cortically damaged patients. *Journal of Abnormal and Social Psychology,* 1958, **57**, 99–106.

DONOHOE, J. W., CURTIN, MARY ELLEN, & LIPTON, L. Interference effects with schizophrenic subjects in the acquisition and retention of verbal materials. *Journal of Abnormal and Social Psychology,* 1961, **62**, 553–558.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin,* 1958, **55**, 316–327.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract No. 8, Nonr–66216, Office of Naval Research.

SPENCE, JANET T. Associative interference on paired-associate lists from extraexperimental learning. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 329–338.

SPENCE, JANET T., & LAIR, C. V. Associative interference in the verbal learning performance of schizophrenics and normals. *Journal of Abnormal and Social Psychology,* 1964, **68**, 204–209.

(Received April 16, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 123–131

SOME CHARACTERISTICS OF POSTHYPNOTIC AMNESIA[1]

JOHN A. WILLIAMSEN, HAROLD J. JOHNSON AND CHARLES W. ERIKSEN
University of Illinois

This experiment investigated (a) the differences in posthypnotic amnesic characteristics of Ss with high and low hypnotic susceptibility and (b) the extent of the amnesia. The experimental Ss were presented 6 words under hypnosis with instructions for amnesia. The simulation Ss pretended they were hypnotized and received the words with instructions for posthypnotic amnesia. The control Ss were given the words with instructions only to remember them. Recognition, recall, and associative tests, administered immediately after, assessed the amnesia. Posthypnotic amnesia impaired recall and recognition among the experimental Ss, but did not reduce the availability of the words as associative responses. The simulating Ss overplayed their amnesic role and also showed impaired performance on the associative tests.

Experiments dealing with posthypnotic amnesia, such as those of Mitchell (1932), Patten (1932), Strickler (1929), and Wells (1941; see Barber, 1962a, for a more extensive review), have shown that the extent of the amnesia obtained is a function of, (*a*) the susceptibility of the subjects used, (*b*) the explicitness of the amnesia suggestions, and (*c*) the face validity or apparent relevance of the assessment procedures to the measurement of amnesia.

These experiments, however, have not been successful in clearly delineating amnesic phenomena. They have not circumscribed the nature and extent of the memory deficits nor, and more importantly, have they succeeded in showing to what extent the amnesic characteristics are unique to the hypnotic state as opposed to other motivating conditions. In particular the posthypnotic-amnesic phenomena have not been differentiated from what would be obtained by nonhypnotized subjects role-playing amnesia or from possible memory differences associated with susceptibility to hypnosis.

It was the purpose of the present experiment to (*a*) determine whether there were quantifiable differences between the amnesic characteristics of subjects possessing high susceptibility and subjects possessing low susceptibility to hypnosis under conditions of hypnosis, simulation, and control; and (*b*) to determine the extent of the amnesia by traditional memory as well as by various associative tasks.

EXPERIMENT I

Method

Subjects. Female volunteers were obtained from an introductory psychology course at the University of Illinois. They were given a form of the Stanford Hypnotic Susceptibility Scale (SHSS; Weitzenhoffer & Hilgard, 1959) adapted for group administration and self-report scoring (Shor & Orne, 1962). Subjects whose score on the scale was 8+ or higher and who passed the posthypnotic amnesia item were classified as high susceptible (HS) and those whose score was 4+ or less and who failed this item as low susceptible (LS).

Group-screening sessions were held until 30 HS and 30 LS subjects were obtained. Because of a positive bias in SHSS scores introduced by the use of a volunteer population (Boucher & Hilgard, 1962) it was necessary to screen about 200 subjects before the LS group was completed. Mean SHSS score for the HS group was 10.5, with the standard deviation = 0.35. For the LS group the mean susceptibility score was 3.0 with standard deviation = 0.33.

Experimental procedure Part I. The 30 subjects in each group were divided evenly into the three experimental treatments, hypnosis, simulation, and control. All subjects were run individually. The subjects assigned to the hypnosis treatment were hypnotized using Form B of the SHSS as part of the induction. Items two through eight of this scale were administered to both the HS and LS subjects so that reliable indications of trance depth could be obtained. The HS subjects passed all seven items as expected and the mean score for the LS subjects was 3.4 with a range of 2–7. Following trance induction, the subject was read a list of words. She was asked to repeat each word and given the suggestion that she would be able to do this and still remain deeply hypnotized. The procedure was

---

[1] This research was supported by Mental Health Grant M-1206 and National Science Foundation Grant G22139.

repeated once. The words (butter, dark, soft, king, slow, cold) were chosen from the Russell-Jenkins Word List (Tinker & Russell, 1958), the criterion for selection being that the first three responses to each word were given by at least 700 of the 1,000 subjects sampled in the Russell-Jenkins study. After the subject had repeated the six words twice she was told:

> Now I want you to listen carefully to what I say. When you awaken you will not be able to remember what the words were. After waking, you will have no memory that I told you these words or what the words were. You will not be able to remember until you hear the signal, "the experiment is over." Then you will be able to remember everything. You will have complete recall of these words after you hear the signal, "the experiment is over," but not before.

The suggestion of amnesia was repeated once more and then the subject was brought out of the trance and led to another room where a second experimenter administered the tests of amnesia.

The subjects in the simulation treatment were told that the experiment was concerned with how well people could simulate posthypnotic amnesia. The procedures administered to the hypnotized group were described and the subject was told:

> Now what I want you to do when you go in to see Mr. Williamsen is to act as if you have been hypnotized and given the suggestion that you could not remember the words. I want you to act as if you had learned the words and then been told that you would have no memory of having learned them until you heard the signal, "the experiment is over." After you hear the signal you will no longer have to act and can tell Mr. Williamsen the words when he asks you. I know you can do a good job. Mr. Williamsen doesn't know if you have been hypnotized or not, so here is a chance to see how good an actress you are. It's not as hard as you think. Now, before I tell you what the words are, do you have any questions?

The subjects were then given the six words in the same manner as the hypnotized group and the simulation instructions were repeated. (Simulators and hypnotized subjects spent the same amount of time with the first experimenter and both groups were given the words about 5 minutes before being taken to another room for the tests of amnesia.)

In the control treatment subjects were given the six words in the same manner as above with the request that they learn them. They were told that the purpose of the experiment would be explained to them by the second experimenter. The control condition served as a check on normal forgetting.

Experimental procedure Part II. A second experimenter administered the memory and associative tasks to enhance the motivation of the simulating subjects (Orne, 1959). Also since the second experimenter was without knowledge as to the experimental treatment administered to the subject, any possible bias on the experimental measures was avoided. The second experimenter asked the hypnotized and simulating subjects if they had been hypnotized. All subjects replied that they had. They were then asked if they were given any instructions by the hypnotist as to how they were to behave when they awoke. Regardless of the answer to this question, subjects were next asked if they had learned any words while hypnotized. Subjects answering affirmatively were further asked to recall as many of the words as they could. This was the first recall test. The following tests of amnesia were then administered in order: partial words, word association, and word recognition.

The partial word test consisted of 12 words printed on 3 × 5 index cards projected on a wall by means of an opaque projector. Six of the words were those learned by the subject (E words); the other 6 were control words (C words) selected from the Russell-Jenkins list according to the same criterion used in the selection of the experimental words. The 6 C words were: long, table, sour, woman, water, and high. The partial words utilized were modeled after the Street-Gestalt figures. Through the progressive deletion of parts of letters the responses to the words can be manipulated. Pretesting of several alternate forms for each of the 12 E and C words resulted in the selection of 12 stimuli which elicited the correct response between 20% and 50% of the time from a group of pilot subjects. Prior to the beginning of the experiment order of presentation was randomized by means of a table of random numbers and this order was held constant over all subjects. The subjects were told that the experimenter was going to show them words that had parts of their letters missing and they were to try to guess what the words were. They were instructed to report their guess as soon as possible, but if subject had not responded in 30 seconds the next partial word was presented. Number of words correctly recognized and response latencies within 30 seconds were recorded.

In the word association test subjects were instructed to respond just as quickly as they could with the first word that came to mind. The stimulus words were the first associations given to the six E and six C words found in the Russell-Jenkins list. Once more order of presentation was randomized and then held constant over all subjects. To further disguise the purpose of this test, an additional six "dummy" words were included in the list. Number of E and C words given as responses and latencies for each association were recorded. Word association was administered twice in succession. In the second administration the subject was instructed to give a response different than the one he had previously given to the stimulus words.

In the word recognition task the six E words, the six C words, and six "dummy" words were typed in random order on a 4 × 6 index card. Subjects were handed the card and asked, "If you had

TABLE 1

MEAN NUMBER OF WORDS RECALLED BY GROUPS ON FIRST AND SECOND RECALL TESTS

Test	Groups					
	High susceptibility			Low susceptibility		
	Hypnotized	Simulator	Control	Hypnotized	Simulator	Control
Recall 1	1.3	0.0	5.4	3.8	0.0	5.5
Recall 2	4.6	5.5	5.4	4.8	4.2	5.6

Note.—Mean differences of 1.0 and 1.1 are significant at approximately the .01 level for within- and between-group comparisons, respectively.

learned any words while hypnotized, are there any words on this card which are words you could have learned?" Number of E words reported by the subject was recorded along with time spent on this task.

For the hypnotized and simulating subjects the signal to remove the amnesia and terminate the experiment was now given and they were requested to give as many of the E words as they could remember (second recall test).

RESULTS

The mean number of items recalled by HS and LS subjects under the three experimental treatments are shown in Table 1 for the first and second recall tests. The overall significance of the differences between these means was evaluated by a modified four-way classification analysis of variance (HS versus LS, treatments, first and second recall, and individuals. Lindquist Type VI). The main effects of treatments ($F = 102.6$ $df = 2/54$) and first and second recall ($F = 236.7$ $df = 1/54$) were both significant beyond the .01 level. In addition the triple interaction between susceptibility, recall test, and treatment ($F = 5.1$ $df = 2/54$) and all three second order interactions were significant beyond the .01 level.

The nature of these significant effects can be seen by turning again to the means in Table 1. To facilitate comparisons between these means Table 1 also contains the mean difference necessary for significance at the .01 level for both within- and between-group comparisons. These values were arrived at by using the within ($s^2 = .7$) and the between subjects ($s^2 = .9$) error terms from the analysis of variance. Since these *t* tests are not precise only the difference for the .01 significance level is used.

Both the LS and the HS simulating subjects differ from both control groups as well as from both hypnotic groups on the first recall test. Also the HS hypnotized subjects differ significantly from both control groups and from the simulators. They also recall significantly fewer words than the LS hypnotized subjects, who in turn show significantly poorer recall than the two control groups. On the second recall measure, this pattern of differences changes and the HS hypnotized subjects no longer differ significantly from either the HS simulators, LS hypnotized or HS and LS controls. In fact, the only group which does differ from any of the other groups is the LS simulating group. These results indicate that hypnotized subjects in both the HS and LS groups show an amnesic effect with regard to recall tests if one test is administered before and another following the removal of a hypnotic suggestion.

The pattern of results among the six subgroups on the recognition memory task presented in Table 2 is very similar to that found on the first recall task. Here the mean number of items recognized is shown along with the mean time the subject spends on the task. The significance of the data in Table 2 was evaluated by two-way classification (LS versus HS and treatments) analyses of variance, one for the number of items recognized and the other for recognition times. For the number of items recognized, both the main effects, LS versus HS ($F = 10.6$ $df = 1/54$) and treatments ($F = 226.4$ $df = 2/54$), were significant beyond the .01 level. The significant effect for susceptibility, however, appears

TABLE 2

Mean Number of Words Recognized and Mean Time in Seconds Spent on Recognition Task by Groups

	Groups					
	High susceptibility			Low susceptibility		
	Hypnotized	Simulator	Control	Hypnotized	Simulator	Control
Words recognized	2.9	0.1	5.5	5.1	0.0	5.5
Time on task	51.9	27.6	46.0	41.4	33.9	44.3

Note.—Mean differences of .43 on words recognized and of 11.7 on time scores are significant at approximately the .01 level.

due to the interaction between susceptibility and treatment ($F = 12.0$ $df = 2/54$) which was also significant. The main effect attributable to susceptibility is no longer significant when tested against the interaction term. This difference indicates that the HS hypnotized subjects recognized fewer words than did the LS hypnotized subjects and further confirms the hypothesis that hypnotic suggestions are more effective with highly susceptible subjects.

There were no significant F ratios in recognition time among the six subgroups. However, the pattern of differences is in keeping with the differences in number of items recognized. Both the HS and LS simulators spent less time on the recognition task relative to the controls while the HS hypnotized spent more time and the LS hypnotized did not differ appreciably from the controls.

The foregoing analyses have shown a significant amnesic effect in the HS hypnotized subjects and a partial amnesic effect in the LS hypnotized subjects. The amnesic effect for the HS hypnotized subjects holds from the first recall through the recognition task and the subjects show recovery of the words on the second recall following the removal of the suggestion. The amnesic effect of the LS hypnotized subjects appears only when the two recall measures are considered. When the recognition measure is considered, the subjects in this group show no amnesia and in fact show a slight decrease from the recognition measure to the second recall. Further, these analyses have shown that the amnesic effect in the HS hypnotized subjects is smaller in magnitude than that shown by simulators from both the HS and LS groups. Also there were no reliable differences between HS and LS subjects in the control condition.

While the words for which amnesia has been suggested are reduced in availability to the hypnotized subjects on direct recall and recognition, there is some question as to

TABLE 3

Mean Number of Partial Word Solutions and Mean Latency in Seconds per Solution by Groups for Experimental and Control Words

	Groups					
	High susceptibility			Low susceptibility		
	Hypnotized	Simulator	Control	Hypnotized	Simulator	Control
Mean solutions						
Experimental words	3.5	1.6	4.6	3.4	1.7	3.9
Control words	0.9	1.9	1.0	1.3	1.2	1.2
Mean latency						
Experimental words	15.7	24.5	10.5	15.7	24.4	15.9
Control words	23.4	22.7	23.8	23.7	23.5	22.4

Note.—For number of solutions mean differences of 1.5 and 1.7 are significant at approximately the .01 level for within- and between-group comparisons, respectively. The corresponding mean differences for the latency data are 1.5 and 6.3.

TABLE 4

MEAN NUMBER OF EXPERIMENTAL AND CONTROL WORDS GIVEN AS ASSOCIATIONS AND MEAN LATENCY IN SECONDS OF ASSOCIATIONS BY GROUPS ON THE TWO ADMINISTRATIONS OF THE WORD ASSOCIATION TEST

	Groups					
	High susceptibility			Low susceptibility		
	Hypnotized	Simulator	Control	Hypnotized	Simulator	Control
Word Association Test 1						
Words						
Experimental words	4.8	1.7	4.6	4.5	4.6	4.0
Control words	3.5	3.6	3.7	3.2	3.7	4.0
Latency						
Experimental words	1.2	2.3	1.2	1.3	1.8	1.0
Control words	1.1	1.7	1.2	1.3	1.7	1.1
Word Association Test 2						
Words						
Experimental words	0.7	0.4	0.5	0.3	0.9	0.7
Control words	1.0	1.1	0.3	0.7	0.6	0.9
Latency						
Experimental words	2.3	2.9	3.1	2.3	2.8	2.1
Control words	2.0	2.3	2.4	2.5	2.6	2.2

whether these words are available to these subjects for use as associative connections and responses. It was to provide an answer to this question that we used the partial words and word association techniques. In Table 3 the mean number of partial word solutions for the experimental and control words is shown for the six subgroups. Table 3 also gives the average latency of the solution to the experimental and control partial words. The significance of these data was again evaluated by analyses of variance. Modified four-way classifications were again employed (Lindquist Type VI). For the number of solutions the only main effect that achieved significance at the .01 level was between experimental and control words ($F = 65.4$ $df = 1/54$). The interaction between E and C words and treatments was also significant at this level ($F = 15.6$ $df = 2/54$).

The analysis of the latencies yielded a significant effect due to experimental and control words ($F = 597.7$ $df = 1/54$), but there the main effect due to treatments was significant ($F = 12.5$ $df = 2/54$). The interactions between words and susceptibility ($F = 17.3$ $df = 1/54$), words and treatments ($F = 239.3$ $df = 2/54$) and words by susceptibility by treatments ($F = 30.7$ $df = 2/54$) were all significant beyond the .01 level when evaluated against the within-error term.

To facilitate comparisons between the mean solutions in Table 3 the difference between the means necessary for significance at the .01 level was again computed using the within-subjects ($s^2 = 1.6$) and the between-subjects ($s^2 = 2.6$) variance estimates from the analyses of variance. With the aid of these values it can be seen in Table 3 that neither the HS hypnotized nor the LS hypnotized subjects differed significantly nor appreciably from either control group on the experimental and control partial words solved. The HS and LS simulators, however, solved significantly fewer than the hypnotized and control groups, but did not differ significantly from each other. A recency effect attributable to exposure to the experimental words during the first part of the experiment is shown by all subgroups except the simulators. This is apparent in that the responses of subjects in the hypnotized and control groups to the experimental partial words yield more correct solutions than their responses to the control words.

A comparison of latency means among the six subgroups parallels the results obtained with the number of correct solutions. Again, it is only the two stimulating groups that show differences from the remaining four subgroups and again the nature of the differences is due to the failure of these subjects to show facilitation for the experimental words relative to the control as a function of recency of experience.

The two administrations of the word association test were scored in terms of the number of experimental and control words given by the subject as responses to the stimulus words. Average latency of response to both experimental and control stimulus words was also recorded. These data are presented in Table 4. Analyses of variance carried out separately on the two administrations failed to yield any effects significant at the .01 level.

The failure of the HS hypnotized subjects to show any effects of the suggested amnesia on the partial words and the word association tests may have been due in part to the masking of differences resulting from the fact that several subjects in this group had shown good recall on the first recall test for the amnesic items. To check this possibility and to further substantiate the finding of no differences on the partial words and word association test attributable to amnesia, a rigid criterion for amnesia on the first recall test was adopted. Seven of the 10 subjects in the HS hypnotized group met a criterion of no more than two of the words recalled on the first recall test plus four additional words recovered after the amnesia was removed. The seven subjects had a mean of 3.3 experimental and .7 control words solved on the partial words test with average latencies of 17.7 for the experimental and 25.6 seconds for the control words. These values are not significantly nor appreciably different from the means for all 10 subjects and do not change the interpretation of the previous analysis on this task. Similarly the results on the two word association tests for these seven subjects failed to change the pattern of results.

Discussion

It is apparent from the data presented that hypnotic trance induction and suggested amnesia produce recall and recognition behavior that is quantitatively different from that observed in highly susceptible subjects who are either simulating amnesia or who have received no instructions regarding amnesia. The HS hypnotized subjects show significantly poorer recall and recognition while under the influence of posthypnotic suggestion relative to control subjects and these differences disappear following removal of the hypnotically induced amnesia. The results with the LS subjects show a similar but attenuated pattern for the recall measures, but show no amnesic behavior on the recognition measure.

The nature of posthypnotic amnesia can be further clarified by considering the results obtained on the less direct measures of memory employed in this study. It is readily apparent from the results of both the word association and the partial words test that suggestions that the subject will not remember the word do not preclude the use of the words as responses. In fact, the hypnotized subjects show a recency effect in their responses to partial words due to the prior exposure to the words during the trance phase of the experiment. This effect is also present in nonhypnotized control subjects, but the subjects who are simulating hypnosis do not show this effect.

The performance of subjects in this study then demonstrates that posthypnotic amnesia varies both with the susceptibility of the subjects and with the type of task used to evaluate the memory deficit. The instructions for amnesia are apparently quite specific in their action and the suggestion that subjects will be amnesic limits the use of the critical words as responses only in situations where it is apparent that memory is being evaluated. The differences between the hypnotized and simulating subjects further indicate that the simulators tend to overplay the role and also that they show a more general inhibition of the use of the critical words as responses in all tasks. These differences in performance between hypnotized subjects and simulators are contrary to several recent findings (Orne, 1959; Sutcliffe, 1960; Barber, 1962b). Perhaps the difference between our results and those of others with respect to simulators is due to the nature of the psychological processes studied. Those studies that failed to find differences between hypnotized and simulating subjects were concerned with a determination of whether subjects can exceed their usual physiological limits under hypnosis on tasks involving pain, muscular strength, and endurance.

Although we have interpreted the findings as indicating that HS hypnotized subjects have shown amnesia in their performance on the various tests, there is an alternative explanation which must be considered. It is possible that the subjects in the hypnosis condition, particularly the HS subjects, did not learn the six experimental words as well as did subjects in the simulating and control groups. This would account for the poorer initial recall performance. The experience these subjects then had with the words on the succeeding tests of partial words and word association gave them an opportunity to increase their learning. Thus the increase in performance on the second recall measure shown by these subjects may not reflect the removal of the amnesia suggestion but be attributable to the increased learning they had acquired while proceeding through the series of tests.

There are several lines of evidence that would argue against this possibility. First are the results obtained from two recent and well-controlled studies that have investigated the question of learning efficiency under hypnosis. Schulman and London (1963) investigated the learning of serial nonsense syllables and of poetry for groups classed as very susceptible, moderately susceptible, and unsusceptible to hypnosis under both hypnotized and nonhypnotized conditions using a counterbalanced design. They found no significant nor suggestive evidence that learning under hypnotized or trance-inducing conditions differed from control conditions. Similarly, Rosenhan and London (1963) found that the recall of nonsense syllables learned under hypnosis did not differ significantly from recall under nonhypnotized conditions for very susceptible subjects. For unsusceptible subjects the hypnosis actually led to a significant increase in recall compared to the nonhypnotized condition.

A second line of evidence is found in the recognition measure in the present study. The recognition test was administered after the first recall, the partial word and the word association procedures had been given, but prior to the removal of the amnesic suggestion. The HS hypnotized subjects showed an appreciable and significant impaired performance on recognition memory. The amnesic suggestion was removed immediately following this recognition measure and the second recall test was then given. At this point the HS hypnotized subjects showed a large gain in recall and their performance did not differ significantly from that of the control subjects. Had these HS hypnotized subjects been making up for original deficits in learning through the various experimental tasks it would seem that this increased learning would have manifested itself on the recognition measure. Also it seems somewhat implausible that within the less than 2-minute period that elapsed from the recognition task to second recall measure interspersed only by the experimenter saying the phrase, "The experiment is over," these subjects would show this gain on direct recall.

While the above lines of evidence seem convincing we nonetheless desired to rule out unequivocally the possibility that our results may have been attributable to less effective learning by the groups in the trance induction procedure. Accordingly, a second control experiment was carried out.

EXPERIMENT II

The purpose of this experiment was twofold. First we wished to determine whether there was less efficient or effective learning in HS subjects when the learning experience occurred under the trance condition obtained in Experiment I. Second we wished to assess the degree of relearning or practice that accrued due to the partial words and the word association procedures that intervened between the first direct recall and the recognition test in Experiment I.

Procedure. Two groups of HS subjects were run. One group (Learning Control) was hypnotized and administered the six experimental words under the identical conditions that pertained for the hypnosis treatment of Experiment I. The same suggestion for posthypnotic amnesia was given and then upon awakening they were administered the partial word and word association procedures. However, these two tests contained only the control words from the previous experiment so that these subjects had no contact with the six experimental words prior to the direct recall measure. Following these procedures the posthypnotic suggestion was removed in the same manner as before and the subject was then administered the first recall test and then the recognition test.

The second group (Practice Control) was administered identical treatment except that upon awakening from the trance they were given a 10 minute waiting period instead of the partial-word and word-association procedures. Following this they were administered in order the direct recall test, the recognition test, the signal for removal of the posthypnotic amnesia, a second recognition test, and a second direct recall measure.

All subjects were hypnotized using Form B of the SHSS and like the HS subjects in Experiment I all subjects passed all items two through eight on this scale.

Subjects. Twenty-one University of Illinois undergraduate female volunteers served as paid subjects. All had been screened by means of group administration of the SHSS Form A scale. All had been selected as highly susceptible to hypnosis on the basis of scores of 10+ or higher. The five subjects assigned to the Learning Control group were directly comparable to the HS subjects in Experiment I. The 16 subjects in the Practice Control group had had a previous trance induction in connection with another experiment and may therefore have been somewhat more practiced in hypnosis than the subjects in the Learning Control group or the subjects in Experiment I.

RESULTS

Table 5 shows the mean recall and recognition scores for the Learning Control group after removal of the posthypnotic amnesia suggestion and for the Practice Control group before and after removal of the amnesia suggestion. The mean recall of 5.2 for the Learning Control subjects does not differ significantly from the mean of 4.6 obtained by the HS hypnotized subjects in Experiment I on the second recall test administered immediately after removal of amnesia ($t = 1.54$, $df = 13$, $p > .10$) nor does it differ significantly from the mean of the HS control group of the preceding experiment ($t < 1.00$). Also the mean recognition score of 5.8 of the Learning Controls after removal of amnesia does not differ significantly from the mean of 5.5 of the HS control group.

TABLE 5

MEAN RECALL AND RECOGNITION SCORES OF LEARNING CONTROL GROUP AFTER REMOVAL OF AMNESIA AND OF THE PRACTICE CONTROL GROUP BEFORE AND AFTER REMOVAL OF AMNESIA

	Groups			
	Learning control[a]		Practice control[b]	
	Recall	Recognition	Recall	Recognition
Before removal of amnesia			.3	.9
After removal of amnesia	5.2	5.8	4.9	5.3

[a] $N = 5$.
[b] $N = 16$.

The recall and recognition scores for the Practice Control group show a marked effect of the removal of the posthypnotic amnesia suggestion. Before removal of the amnesia they recall and recognize significantly fewer words than did the HS hypnotized subjects in the preceding experiment ($t = 2.38$; 12.50, respectively for recall and recognition, with $p < .05$). After removal of the amnesia these Practice Control subjects do not differ significantly on either recall or recognition from the HS control subjects of Experiment I nor from the HS hypnotized subjects on their second recall.

DISCUSSION

The results of our second experiment seem quite clear in demonstrating that the performance of the HS hypnotized subjects in Experiment I cannot be attributed to deficiency in original learning of the six experimental words. The Learning Control subjects who had no recall tests nor contact with the experimental words while under the posthypnotic amnesia suggestion, had recall and recognition scores directly comparable to the HS controls when the amnesia suggestion was removed.

The results from the Practice Control subjects seem equally clear in demonstrating little or no effect attributable to contact with the experimental words in the form of the partial-words or word-association procedures. These subjects were not administered these procedures and yet their recall after removal of the amnesia was slightly better, though not significantly, than that of the HS hypnotized subjects on their second recall.

The Practice Control subjects before removal of the amnesia recalled and recognized significantly fewer words than did the HS hypnotized subjects under the amnesia suggestion. This is probably attributable to the somewhat greater experience in hypnosis of the Practice Control subjects. These latter subjects had all participated in another hypnosis experiment prior to serving in the present study.

In summary, the present experiments have shown that suggested amnesia varies with both susceptibility and the task used to assess the degree of amnesia. The HS hypnotized subjects differ from the control subjects in terms of the amount of recall and they also differ from simulating subjects who overplay the amnesic role. While the present studies do not answer all of the questions concerning posthypnotic amnesia, they have demonstrated that such amnesia can be produced under controlled conditions and have suggested some of the variables which influence its expression.

REFERENCES

BARBER, T. X. Toward a theory of hypnosis: Posthypnotic behavior. *A.M.A. Archives of General Psychiatry,* 1962, **7**, 321–342. (a)

BARBER, T. X. Experimental controls and the phenomena of "hypnosis"; A critique of hypnotic research methodology. *Journal of Nervous and Mental Disease,* 1962, **134**, 439–505. (b)

BOUCHER, R. G., & HILGARD, E. R. Volunteer bias in hypnotic experimentation. *American Journal of Clinical Hypnosis,* 1962, **5**, 49–51.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MITCHELL, M. B. Retroactive inhibition and hypnosis. *Journal of General Psychology,* 1932, **7**, 343–358.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58**, 277–299.

PATTEN, E. F. Does post-hypnotic amnesia apply to practice effects? *Journal of General Psychology,* 1932, **7**, 196–201.

ROSENHAN, D., & LONDON, P. Hypnosis in the unhypnotizable: A study in rote learning. *Journal of Experimental Psychology,* 1963, **65**, 30–34.

SCHULMAN, R. E., & LONDON, P. Hypnosis and verbal learning. *Journal of Abnormal and Social Psychology,* 1963, **67**, 363–370.

SHOR, R. E., & ORNE, EMILY F. *The Harvard Group Scale of Hypnotic Susceptibility, Form A: An adaptation for group administration with self-report scoring of the Stanford Hypnotic Susceptibility Scale, Form A.* Cambridge: President & Fellows of Harvard College, 1962.

STRICKLER, C. B. A quantitative study of posthypnotic amnesia. *Journal of Abnormal and Social Psychology,* 1929, **24**, 108–119.

SUTCLIFFE, J. P. "Credulous" and "skeptical" views of hypnotic phenomena: A review of certain evidence and methodology. *International Journal of Clinical and Experimental Psychology,* 1960, **8**, 73–101.

TINKER, M. A., & RUSSELL, W. A. *Introduction to methods in experimental psychology.* (3rd ed.) New York: Appleton-Century-Crofts, 1958.

WEITZENHOFFER, A. M., & HILGARD, E. R. *The Stanford Hypnotic Susceptibility Scale.* Palo Alto: Consulting Psychologists Press, 1959.

WELLS, W. R. The extent and duration of posthypnotic amnesia. *Journal of Psychology,* 1941, **9**, 137–151.

(Received May 1, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 2, 132–154

EXPERIMENTAL ANALYSES OF "HYPNOTIC" BEHAVIOR:

A REVIEW OF RECENT EMPIRICAL FINDINGS[1]

THEODORE XENOPHON BARBER

Medfield Foundation, Harding, Massachusetts

Experiments are reviewed that were designed to delineate the instrumental variables in eliciting objective and subjective responses to test suggestions of body immobility, analgesia, hallucination, deafness, selective amnesia, and so on, that is, in eliciting behaviors of the type traditionally termed "hypnotic." The question at the forefront of discussion is: In producing "hypnotic-like" behavior, what are the relative effects of (a) S's personality characteristics, and of instructional-situational variables, such as (b) defining the situation to S as "hypnosis" or "control," (c) administering task-motivational instructions, (d) administering suggestions of relaxation, drowsiness, and sleep, and (e) suggesting to S that he can now easily respond to test suggestions?

Weitzenhoffer and Sjoberg (1961) and Barber and Glass (1962) recently demonstrated experimentally that "hypnotic induction procedures" are effective, as compared to an uninstructed control condition, in facilitating objective and subjective responses to test suggestions of the type comprising the Stanford Hypnotic Susceptibility Scale and the Barber Suggestibility Scale (BSS), e.g., test suggestions of limb rigidity, body immobility, hallucination, and amnesia. However, the hypnotic induction procedures used in these experiments, and in most other recent experiments, included at least four distinct independent variables: (*a*) the situation was defined to the subject as hypnosis; (*b*) motivational instructions were administered e.g.,

> in participating here you are contributing to scientific knowledge. . . . All I ask of you is that you keep up your attention and interest and continue to cooperate. . . . Most people find this a very interesting experience . . . [Weitzenhoffer & Hilgard, 1959, pp. 9, 14].

(*c*) repeated suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, and sleep were administered; finally, (*d*) it was suggested to the subject that he could now easily respond to test suggestions and could easily experience the suggested effects e.g.,

> You will be able to do all sorts of things I ask you to do. . . . You are going to experience many things I will tell you to experience . . . [p. 16].

It should be noted that the subjects participating in the Weitzenhoffer and Sjoberg and in the Barber and Glass experiments received motivational instructions under the hypnotic induction condition but not under the control condition.

The first question posed in the present review is: Suppose an attempt had been made to motivate the controls to perform to the best of their ability on experimental tasks. Would such a motivated group have been as responsive, overtly and subjectively, or less responsive to the test suggestions than the group that received the hypnotic induction? After an answer to this question is provided, the review focuses on the four independent variables that are typically included in present-day hypnotic inductions, and asks: are each of these variables effective, or are some ineffective, in facilitating response to test suggestions? What is the relative contribution of each effective variable? Do the variables interact to facilitate suggestibility? Upon summarizing a series of experiments that was designed to answer these questions, a final query is posed: what are the effects of personality factors and of several relatively subtle instructional-situational factors, such as implicit suggestions, that the procedure to be employed is effective in inducing hypnosis?

---

[1] The writing of this paper was made possible by a research grant (MH-07003) from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service.

Since the majority of experiments pertinent to the above questions were conducted in the writer's laboratory, the present paper constitutes primarily a summary of recent research by the writer and his collaborators. Focusing on a limited set of questions, no attempt is made to review all recent experimental work concerned with hypnotic behavior. For broad coverage of other recent studies in this area, which asked different questions than those asked here, the reader is referred to Barber (1961a, 1962a, 1962b, 1964b, 1964e, 1965b), Gordon (in press), Hilgard (in press), and Schneck (1963).

PART I. THE SUGGESTIBILITY-ENHANCING EFFECTS OF "HYPNOTIC INDUCTION PROCEDURE" AND OF "TASK MOTIVATIONAL INSTRUCTIONS"

A series of recent experiments were designed (*a*) to ascertain whether brief instructions intended to motivate the subject to perform well in the test situation ("task motivational instructions") are effective in facilitating response to test suggestions, and (*b*) to compare the effects on suggestibility of task motivational instructions and of a standardized procedure of the type commonly termed a hypnotic induction. The dependent variables in these experiments included response to the BSS plus response to suggestions of auditory and visual hallucination, enhanced strength and endurance, enhanced learning-cognitive proficiency, analgesia, and deafness. Experiments concerned with these and other dependent variables are reviewed in turn.

Barber Suggestibility Scale

Barber and Calverley (1962, 1963b, 1963c) carried out a series of similar experiments, each of which included at least three treatment groups. One group received a standardized 15-minute hypnotic induction procedure. This procedure, which was patterned after the induction procedures of Friedlander and Sarbin (1938), Marcuse (1959, p. 53), and Weitzenhoffer and Hilgard (1959), included the four variables listed in Table 1. The experimental treatment for the second group consisted of task motivational instructions, administered for 60 seconds, designed to produce positive motivation to perform well in a situation defined to the subject as "a test of imagination":

TABLE 1

INDEPENDENT VARIABLES TYPICALLY INCLUDED IN PRESENT-DAY "HYPNOTIC INDUCTION PROCEDURES"

1. The situation is defined to the subject as "hypnosis."
2. Instructions are administered that are designed to motivate the subject to give a good performance.
3. Suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, and sleep are administered.
4. It is suggested to the subject that he can now easily respond to test suggestions and can easily experience suggested effects.

> I'm going to test your ability to imagine and to visualize. . . . Everyone passed these tests when they tried. . . . What I ask is your cooperation in helping this experiment by trying to imagine vividly what I describe to you. . . . If you try to imagine to the best of your ability, you can easily imagine and do the interesting things I tell you. . . .

A third group (Control) was also told that it was to receive a test of imagination but this group received neither a hypnotic induction nor task motivational instructions. All groups were tested on objective and subjective responses to the BSS which includes eight standardized test suggestions. (Since the BSS will be referred to often below, it is summarized for convenience of reference in Table 2.) Each of the experiments in this series yielded essentially the same results: (*a*) the group given the standardized 15-minute hypnotic induction procedure and the group given the 1-minute task motivational instructions were significantly more responsive to the BSS than the control group; and (*b*) the hypnotic induction group and the task motivational group did not differ significantly from each other on either objective or subjective responses to the BSS.

In summary, the three experiments in this series indicated that brief task motivational instructions are about as effective as a standardized procedure of the type traditionally labeled as a hypnotic induction in facilitating objective and subjective responses to suggestions of arm levitation, thirst "hallucination,"

TABLE 2

BARBER SUGGESTIBILITY SCALE

Test Suggestions

1. Arm Lowering

"Imagine that your right arm is feeling heavier and heavier . . . it's moving down and down . . . [Barber & Calverley, 1963a, p. 590]."

2. Arm Levitation

"Imagine that the [left] arm is becoming lighter and lighter . . . it's moving up and up . . . [p. 590]."

3. Hand Lock

"Imagine that your [clasped] hands are two pieces of steel that are welded together . . . it's impossible to pull your hands apart. . . . Try, you can't . . . [p. 590]."

4. Thirst "Hallucination"

"Imagine that you've just finished a long, long walk in the hot sun. . . . You feel thirstier and thirstier, dry and thirsty . . . [p. 591]."

5. Verbal Inhibition

"Imagine that the muscles in your throat and jaw are solid and rigid. . . . The harder you try to say your name the harder it will become. Try, you can't . . . [p. 591]."

6. Body Immobility

"Imagine that you've been sitting in that chair so long that you're stuck to it . . . stuck in the chair; you can't stand up . . . try, you can't . . . [p. 591]."

7. "Posthypnotic-Like" Response

"When this experiment is over . . . I'll click like this and you'll cough automatically . . . [p. 591]."

8. Selective Amnesia

"You'll remember all the tests . . . all except for . . . the one where I said your arm was becoming lighter and moving up . . . you will not remember until I [say] 'Now you can remember'. . . [p. 591]."

Scoring Criteria

Objective scores: Criteria used to score responses objectively (together with norms and reliability data) are presented in Barber and Glass (1962), Barber and Calverley (1963a), and Barber, Karacan, and Calverley (1964). The maximum objective score obtainable on the scale is eight points (one point for each test suggestion passed).

Subjective scores: Immediately after the scale has been administered and scored objectively, the experimenter describes each of the test suggestions that the subject has passed with an objective score of either one half of one point and asks, "Did you actually feel [the suggested effect] or did you go along with the suggestion in order to follow instructions or to please me?" Subjective scores are assigned as follows: one point for each test suggestion that the subject states that he "actually felt."

---

body immobility, inability to speak one's name, selective amnesia, and the other test suggestions described in Table 2. This finding gives rise to an important question: Are task motivational instructions also effective in facilitating response to test suggestions other than those included in the BSS (for instance, suggestions intended to produce auditory and visual hallucinations)? If task motivational instructions are effective in facilitating responses to such suggestions, how effective are they as compared with a hypnotic induction procedure?

Auditory and Visual Hallucinations

A series of clinical and experimental studies, reviewed elsewhere (Barber, 1964f) indicated that auditory and visual hallucinations can at times be evoked from subjects who have received a hypnotic induction by suggesting to them that they shall hear sounds and see objects that are not present. Until recently, however, no one had attempted to ascertain whether similar hallucinatory behavior can be elicited, without a hypnotic induction, by instructions designed to motivate subjects to try to hear and to see nonpresent stimuli.

Barber and Calverley (1964a) included this comparison in a recent experiment, described below.

Seventy-eight female volunteers were individually pretested as follows: (*a*) Immediately upon being seated in the experimental room the subject was informed that the experiment would begin at once and was told, firmly and seriously, to close her eyes and to hear a phonograph record playing a specified tune. (*b*) After 30 seconds she was asked to open her eyes and to check a standardized rating scale regarding the vividness and realism of the suggested auditory hallucination. (*c*) The subject was next given suggestions, in a firm tone of voice, to see a cat sitting on her lap and, finally, (*d*) she was asked to check a second rating scale with regard to the vividness and reality of the suggested visual hallucination.

In the same experimental session, each subject was retested individually on equivalent hallucination suggestions (to hear a record playing another tune and to see a dog sitting on her lap) after the administration of one of three experimental treatments. Twenty-six subjects were randomly assigned to each treatment group. Subjects allocated to Group 1 were given the suggestions to hallucinate after receiving a standardized 15-minute hypnotic induction procedure (which included the variables listed in Table 1). The subjects assigned to Group 2 were given the identical suggestions to hallucinate after receiving task motivational instructions ("This time I want you to really try to see and to hear the things I ask you to. . . . Don't assume that it can't be done. It's really quite easy . . .)." Subjects assigned to Group 3 (Control) received the identical suggestions to hallucinate that were given to Groups 1 and 2, but without receiving a hypnotic induction or task motivational instructions. Analyses of variance and covariance showed that Groups 1 and 2 did not differ significantly from each other but both groups were significantly more responsive to the suggestions to hallucinate than Group 3. That is, response to the hallucination suggestions was enhanced to a comparable degree (above the subjects' own pretest level and also above the retest level manifested by the Control group) by the hypnotic induction procedure and also by the task motivational instructions.

Enhanced Strength and Endurance

Several studies (e.g., Hadfield, 1924; Ikai & Steinhaus, 1961; Wells, 1947) indicate that suggestions of enhanced strength and endurance, given subsequent to a hypnotic induction procedure, tend to facilitate performance on physical tasks. These studies, however, did not include a comparison group given similar suggestions of heightened strength and endurance without a hypnotic induction. Orne (1959), Levitt and Brady (1964), London and Fuhrer (1961), and Barber and Calverley (1964l) have recently conducted experimental studies that included this required comparison group.

Orne (1959) worked with nine selected "hypnotizable" male subjects. Each was tested on a weight-holding endurance task, first under a hypnotic induction condition and then under a task motivational condition. Under the hypnotic induction condition, suggestions were given that the subject would not experience pain or fatigue and would manifest a high level of performance. Under the task motivational condition the subjects were told that females are able to hold the weight for a certain period of time (which was equal to the subject's performance under the hypnotic induction condition), and that the subject would receive a small monetary reward if he surpassed the performance typical of females. Seven of the nine subjects manifested increased endurance under the task motivational condition. When the remaining two subjects were given a second chance with additional task motivational instructions, they too exceeded the performance they had manifested under the hypnotic induction condition. Although this study suggested the possibility that task motivational instructions given without hypnotic induction may be more effective than suggestions for high performance given subsequent to a hypnotic induction, it did not conclusively demonstrate this point. The experimental conditions were not counterbalanced and the possibility was not excluded that the higher level of performance manifested under the second treatment (task motivation) may have been due to familiarity with or practice on the task.

Levitt and Brady (1964) compared weight-holding endurance in eight female subjects under three counterbalanced experimental conditions: (*a*) a hypnotic induction condition; (*b*) a hypnotic induction condition that included suggestions of anesthesia of the arm and shoulder; and (*c*) a task motivational condition comprised of verbal exhortation plus monetary reward for high performance. Weight-holding endurance did not differ significantly under the three conditions.

London and Fuhrer (1961) worked with 32 female subjects. In the first session, the subjects were tested under counterbalanced hypnotic induction and nonhypnotic induction conditions for grip strength and weight-holding endurance. In the second session, they were retested on the same tasks under counterbalanced hypnotic induction and nonhypnotic induction conditions with task motivational instructions administered under both conditions: "give it everything you've got . . . make this a total effort . . . let me see how really far you can go!" Results were: hypnotic induction, without task motivational instructions, did not affect either strength or endurance; task motivational instructions, given with and also without a preceding hypnotic induction, facilitated strength and endurance. London and Fuhrer concluded that "hypnosis as such adds nothing magical to performance. Motivational instructions are clearly more important than hypnotic trance."

Barber and Calverley (1964l) conducted an experiment along similar lines. Sixty female volunteers were first individually pretested on a grip strength and a weight-holding endurance task and then were retested on the same tasks under one of four experimental treatments, with 15 subjects assigned at random to each treatment. The experimental treatment for one group (task motivation) consisted of instructions designed to produce positive motivation to perform maximally: "It is possible for everyone to increase their strength tremendously when they set their mind upon doing much better. . . ." A second group was given a hypnotic induction procedure, which included the variables listed in Table 1, followed by suggestions of enhanced strength and endurance. A third group received a hypnotic induction but did not receive task motivational instructions. A fourth group (Control) was retested without special instructions or suggestions. With respect to the grip strength task, analyses of variance and covariance failed to show significant differences among the four groups. With respect to the weight-holding endurance task, subjects receiving the hypnotic induction *without* suggestions of enhanced endurance showed a decrement in performance on retest as compared to pretest. However, subjects receiving the hypnotic induction together with suggestions of enhanced endurance, and also subjects receiving task motivational instructions without a hypnotic induction, manifested improved performance on retest as compared to their own pretest performance and also as compared to the retest performance of the control group.

In brief, the above experiments appear to be in line with previous studies, reviewed in detail elsewhere (Barber, in press), which indicated that suggestions of enhanced strength and endurance given subsequent to a hypnotic induction procedure tend to facilitate performance. In addition, the above experiments extend previous studies in indicating that whatever improvement in performance can be produced by suggestions of enhanced strength and endurance given under a hypnotic induction condition can also be produced by task motivational instructions given without a preceding hypnotic induction.

Enhanced Learning-Cognitive Proficiency

A series of experiments (e.g., Hammer, 1954; Illovsky, 1963; Sears, 1955), recently reviewed elsewhere (Barber, 1965a), indicated that suggestions of enhanced learning-cognitive proficiency, given subsequent to a hypnotic induction procedure, tend to facilitate performance on some, but not all, intellectual tasks. A crucial comparison condition was not included in these experiments: suggestions for enhanced learning or cognitive proficiency given without a hypnotic induction. Salzberg (1960a, 1960b), Fowler (1961), and Parker and Barber (1964) recently conducted experiments that included this comparison.

Salzberg (1960a, 1960b) tested three experimental groups on one simple and two relatively complex intellectual tasks (simple counting, memory for nouns, and abstract reasoning). Group 1 received a hypnotic induction procedure and suggestions for high performance (task motivational suggestions). Group 2 was treated in the same way as Group 1 with the exception that subjects in Group 2 performed the tasks after they were told to "awaken" from hypnosis. Group 3 received the identical task motivational suggestions that were given to Groups 1 and 2 without a hypnotic induction. In general, Groups 1 and 2 performed more proficiently on the tasks than Group 3. However, the subjects had not been randomly assigned to the three experimental groups. Subjects allocated to Groups 1 and 2, but not to Group 3, were preselected for "hypnotizability." Operationally, criteria for hypnotizability consisted of positive responses to test suggestions, e.g. suggestions of body sway and hand levitation. By thus confounding the experimental treatments with pre-existing differences among subjects with respect to hypnotizability or "suggestibility," Salzberg failed to ascertain whether Groups 1 and 2 were more responsive than Group 3 to task motivational suggestions because they (*a*) received the suggestions after they had received a hypnotic induction or (*b*) were more responsive to suggestions initially.

Fowler (1961) worked with 40 selected hypnotizable students. Each was pretested on Form I of the Iowa Silent Reading test and the Otis Mental Ability test. The subjects were then randomly assigned to two groups and retested on Form II. One group was given a hypnotic induction together with task motivational instructions of enhanced concentration, comprehension, memory, and accuracy. The other group was given similar task motivational instructions without a hypnotic induction. Both groups manifested improved retest performance on the reading test but not on the mental ability test. The two groups did not differ significantly on the degree of improvement shown on the reading test. Although this experiment seemed to indicate that task motivational instructions facilitate reading performance when given with and also without a hypnotic induction, it failed to demonstrate this point conclusively. Because a control group was not retested without task motivational instructions, the possibility was not excluded that improved retest performance was due to familiarity with or practice on the task.

Parker and Barber (1964) pretested their subjects on three tasks: Digit Symbol Substitution adapted from the Wechsler-Bellevue; the Abstract Reasoning task from the Differential Aptitude Test Battery; and a Memory for Words task. Retests on equivalent forms of the tasks were given under the following three experimental treatments with 10 "suggestible" subjects (each had obtained relatively high scores on the BSS) randomly assigned to each treatment. Treatment 1 included only task motivational instructions: "If you try real hard, you will do better. Everyone who has tried harder on these tests has done better the second time and I'm sure you can too. . . ." Treatment 2 consisted of a hypnotic induction procedure followed by task motivational instructions. Treatment 3 (Control) did not include either the hypnotic induction or the task motivational instructions. A fourth group of 10 nonsuggestible subjects (who had previously obtained relatively low scores on the BSS), was retested under Treatment 1 (task motivational instructions). Analyses of covariance showed that the retest performance of the four groups did not differ significantly on the Memory for Words or the Abstract Reasoning tasks. However, subjects under Treatments 1 and 2 showed significantly enhanced retest performance on the Digit Symbol Substitution task as compared to the controls, that is, the task motivational instructions produced equivalent enhancement of performance in both suggestible and nonsuggestible subjects. Similarly, those who had received and those who had not received the hypnotic induction showed equivalently enhanced performance after task motivational instructions.

Reduced Pain Reactivity ("Analgesia")

In a recent experiment by Barber and Hahn (1962), the dependent variables were subjective responses and physiological responses to a pain-producing stimulus (immersion of a hand in freezing water, at 2 degrees centigrade, for 3 minutes). Forty-eight subjects who had been previously rated as suggestible (obtaining relatively high scores on the BSS) were assigned at random to four treatment groups with 12 subjects per group. One group received a standardized hypnotic induction procedure (which included the four variables listed in Table 1) followed by suggestions of hand anesthesia administered for 1 minute: "The hand is losing all feeling, all sensation, it is becoming more and more numb, you will be unable to feel anything at all with it. . . ." Upon completion of the anesthesia suggestions, the hand was exposed to the pain-producing stimulus. A second experimental group (task motivation) received an experimental treatment consisting of 1-minute instructions designed to motivate the subjects to imagine a pleasant situation when exposed to the noxious stimulation—e.g., "If you try to imagine that the water is pleasant and try to think of it as *not* uncomfortable, you will be able to keep your hand in the water without being bothered by it at all. . . . I'm sure that you'll be able to continue imagining this and that you will not fail the test." Two additional groups did not receive special instructions: one group (Uninstructed) was exposed to the noxious stimulus and the other (Control) was exposed to a nonpainful stimulus (immersion of the hand in water at room temperature). The group receiving the hypnotic induction and the group receiving the task motivational instructions did not differ significantly in either subjective reports of pain experienced or in four physiological responses to the noxious stimulus that were monitored by a polygraph (heart rate, forehead muscle tension, respiratory irregularities, and skin resistance). As compared to the uninstructed treatment, both the hypnotic induction and the task motivational treatments were effective in attenuating pain reactivity, as indicated by subjective reports and by reduction in respiratory irregularities and forehead muscle tension. However, under both the hypnotic induction and the task motivational treatments, subjects reported more discomfort and pain and showed faster heart rate and lower skin resistance than the control subjects who received the nonpainful stimulus. In brief, the results indicated that subjective responses and physiological responses to a pain-producing stimulus can be reduced to a comparable degree (but are not abolished) by (*a*) administering suggestions of anesthesia immediately upon completion of a hypnotic induction procedure or by (*b*) administering instructions, without a preceding hypnotic induction, designed to motivate the subject to imagine vividly a pleasant situation when exposed to noxious stimulation.

The pain-reducing effect obtained in the above experiment with the hypnotic induction treatment is consistent with previous studies in this area, reviewed elsewhere (Barber, 1963). The effects obtained with the task motivational treatment are generally in line with previous studies (Brown & Vogel, 1938; Shor, 1962; Sutcliffe, 1961) which failed to find significant differences in physiological responses to pain-producing stimuli under a "hypnotic anesthesia" treatment and a "waking" treatment that included instructions to try not to react to painful stimulation.

Suggested Deafness

A series of experimental studies (e.g., Erickson, 1938a, 1938b; Kline, Guze, & Haggerty, 1954; Malmo, Boag, & Raginsky, 1954; Pattie, 1950; Sutcliffe, 1961), reviewed in detail elsewhere (Barber, 1964c), demonstrated that suggestions of deafness given subsequent to a hypnotic induction procedure are at times effective in eliciting subjective reports that auditory stimuli either were not perceived at all or were perceived as reduced in intensity. In each of these experiments, suggestions of deafness were given together with a hypnotic induction, but similar suggestions of deafness were not given without a hypnotic induction. Barber and Calverley (1964f) included this needed comparison condition in an experiment designed to answer three questions: (*a*) Are suggestions of deafness effective in producing partial or total "subjective deafness" when given immediately upon completion of a "hypnotic induction?" (*b*) Are suggestions of deafness also effective when given to a group that has received only task motivational instructions? (*c*) Does subjective deafness produce objective consequences that are similar to, or identical with, the objective effects of actual deafness?

Delayed auditory feedback, by which the subject's utterances are delayed for a fraction of a second, amplified, and returned to his ears through a headphone, was used to assess the objective consequences of subjective deafness. (Previous studies, reviewed by Yates, 1963, had shown that, in the individual with normal hearing, delayed auditory feedback produces speech disturbances such as stuttering, mispronunciations, increased vocal intensity, and slowed rate of speaking.) Forty-two women college students were given a pretest on Form A of the Gilmore Oral Reading test with delayed feedback. Retests on Form B, also with delayed feedback, were given under three experimental treatments, with 14 subjects assigned at random to each treatment. Subjects allocated to Treatment 1 received a standardized 15-minute hypnotic induction procedure (which included the variables listed in Table 1) followed by suggestions of deafness ("you will be deaf to your voice, you will not hear your voice coming back to you, you will not hear anything at all . . ."). Subjects allocated to Treatment 2 were given "task motivational instructions" together with deafness suggestions ("I want you to ignore and be deaf for your voice coming back to you . . . you must try hard, try very hard. It can be done, so please, I do not want you to disappoint me . . ."). Subjects allocated to Treatment 3 (Control) were retested without receiving either a hypnotic induction or deafness suggestions. Significantly more subjects reported "partial deafness" (sounds were perceived as reduced in intensity) or "total deafness" (sounds were not heard at all) under Treatment 2 (93%) than under Treatment 1 (64%). Subjects under both Treatment 1 and 2 who claimed partial or total "deafness" responded to the delayed auditory feedback of their speech in the same way as subjects under Treatment 3 (Control) who heard normally, that is, with stuttering, mispronunciations, slowed rate of speaking, and increased vocal intensity. In brief, this experiment indicated that: (*a*) subjective reports of partial or total deafness can be elicited more effectively by simply instructing subjects under a "waking" experimental treatment to try to the best of their ability not to hear rather than by administering suggestions of deafness under a hypnotic experimental treatment; and (*b*) the objective effects of partial or total "suggested deafness," as evaluated by the method of delayed auditory feedback, closely resemble the objective effects of normal hearing and appear to be markedly dissimilar to the effects of actual deafness. Further studies are clearly needed to confirm these striking experimental results.

Suggested Salivation

The dependent variables in most of the above experiments consisted of subjective responses and overt behavioral responses to test suggestions. A recent experiment by Barber, Chauncey, and Winer (1964) carried out along similar lines as those described above, used a physiological response—salivary secretion to suggested stimuli—as the dependent variable. This experiment set out to confirm and extend previous studies (Bowles & Pronko, 1949; Pronko & Hill, 1949) which indicated that, in subjects who have received a hypnotic induction, suggested taste stimuli tend to affect salivary secretions in a similar manner as actual gustatory stimuli, e.g., suggestions that an acid solution is tasteless tend to depress the salivary secretory response to an acid solution placed on the tongue, and suggestions that water has a sour taste tend to elevate the salivary response to water. Barber et al. (1964) asked: (*a*) Are suggestions designed to produce positive motivation to imagine vividly that an acidic solution is tasteless or that water is sour effective in depressing and elevating salivary secretions, respectively? (*b*) Are such suggestions more effective when given with, or when given without, a preceding hypnotic induction procedure? Under counterbalanced hypnotic induction and nonhypnotic induction conditions, 16 subjects were given brief instructions designed to produce positive motivation to imagine vividly the suggested taste stimuli ("Disregard the actual taste of the solution we are using today, and vividly imagine that [an acidic solution is tasteless or that water is sour] . . . continue to imagine vividly . . . to the very best of your ability . . ."). The results indicated that: the parotid gland salivary response to the acidic solution was depressed by the suggestion that the solution was tasteless whereas the response to water was enhanced by the suggestion that the water was sour; and the effects of suggestions on salivation were not significantly different under the hypnotic induction and the nonhypnotic induction conditions.

Colorblindness, Amnesia, Imagined Pain Stimulation, Dreaming on a Specified Topic, and Time Distortion

Other recent experiments, carried out along similar lines as those described above, appear to indicate the following:

1. Very similar responses to the Ishihara Colorblindness test can be elicited either by (*a*) administering a hypnotic induction and suggestions of red-, green-, or total-colorblindness or by (*b*) instructing subjects who have not received a hypnotic induction to try to ignore the colors red or green or to try to respond as if they are colorblind (Barber, 1964c; Barber & Deeley, 1961; Bravin, 1959; Rock & Shipley, 1961).

2. Suggestions to forget preceding experimental events are as effective, if not more effective, with "awake" subjects than with hypnotized subjects in producing subjective reports of amnesia (Barber & Calverley, 1962, 1964m).

3. Several physiological responses that are characteristically evoked by a cold-pain stimulus (increased heart rate and forehead muscle tension and a tendency toward reduction in skin resistance) can be elicited with or without a preceding hypnotic induction by instructions designed to motivate the subject to try to imagine vividly the cold and painful stimulus (Barber & Hahn, 1964). However, contrary to the general trend of the experiments reviewed above, instructions to imagine the cold-pain stimulus were more effective in this experiment in eliciting subjective reports of discomfort and pain when given with, rather than without, a hypnotic induction.

4. Direct suggestions to dream the following night on a specified topic appear to be about as effective when given without as when given with a preceding "hypnotic induction" in eliciting reports of having dreamed on the suggested topic (Barber, 1962d; Barber & Calverley, 1962; Fisher, 1953).

5. Subjective reports indicative of "time distortion" can be elicited from the majority of volunteer experimental subjects, with or without a preceding hypnotic induction, by suggestions that a brief period will seem to be a long period of time (Barber & Calverley, 1964j).

Theoretical Considerations

By and large the experiments summarized above appear to indicate two interrelated conclusions: (*a*) Task motivational instructions, given with or without a preceding hypnotic induction procedure, tend to facilitate performance on tests of strength and endurance, tend to produce enhanced proficiency on some intellectual tasks (such as Digit Symbol Substitution), and tend to facilitate the elicitation of several singular physiological effects (such as salivary secretions to imagined gustatory stimuli). (*b*) Response to a variety of test suggestions—e.g., response to the BSS, and to suggestions of deafness, visual and auditory hallucinations, and reduced pain reactivity—can be facilitated by two sets of variables: one set subsumed under the label hypnotic induction procedure and the other set subsumed under the label task motivational instructions. The latter conclusion raises an important question for theories of hypnotic behavior: do these two sets of variables facilitate response to test suggestions by similar or by dissimilar processes?

It may be that dissimilar intervening processes are involved. For instance, it may be that procedures of the type labeled as hypnotic inductions (as outlined in Table 1) produce heightened suggestibility by giving rise to a hypnotic state or "trance state," whereas task motivational instructions facilitate suggestibility by other processes that do not involve a "hypnotic trance state." On the other hand, it may be that both hypnotic induction procedures and task motivational instructions facilitate suggestibility by giving rise to a hypnotic state. It could also be, of course, that both hypnotic induction procedures and task motivational instructions produce enhanced suggestibility by processes that do *not* involve a trance state. These hypotheses are difficult if not impossible to prove or disprove at the present time. Criteria for denoting the presence or absence of the hypnotic state or trance are not only ambiguous but also differ among investigators and the same investigator sometimes uses different criteria at different times. Researchers have sought a physiological index of the trance state or the hypnotic state for nearly a century but have failed to find one (Barber, 1961b, 1964e). Since a physiological index is not available, present-day investigators generally use one or both of the following behavioral criteria to denote the presence of hypnosis or hypnotic trance:

1. At times investigators categorize a subject as being in hypnosis or trance if he shows a high level of response to test suggestions of the type included in the Stanford Hypnotic Susceptibility Scale or the BSS, e.g., test suggestions of body immobility, hallucinations, and amnesia. When the term is used in this way the supposed antecedent variable (hypnotic trance) is inferred from the consequent or dependent variable (positive response to test suggestions). This usage is tautological: a person is said to respond to test suggestions because he is in hypnotic trance and he is said to be in hypnotic trance because he responds to test suggestions. (Denoting the hypnotic state operationally, by referring to the administration of a hypnotic induction procedure is, of course, also tautological.)

2. At other times a subject is said to be hypnotized or to be in the hypnotic state or in trance if he manifests such signs or characteristics as rigid facial expression, disinclination to talk, "literal-mindedness," and lack of spontaneity and initiative (Erickson, Hershman, & Secter, 1961, pp. 55–58; Ludwig & Lyle, 1964; Pattie, 1956, p. 21; Weitzenhoffer, 1957, p. 212). Difficulties arise when hypnotic trance is inferred from these characteristics and is then used as a construct to explain high response to test suggestions of body immobility, hallucination, amnesia, and so on. For instance, these supposed signs of trance may themselves be responses to suggestions of relaxation, drowsiness, sleep, and passivity given during the "induction" procedure. (If this is the case, then all that is being said is that subjects who are responsive to certain types of suggestions, e.g., suggestions of drowsiness and passivity, are also responsive to other types of suggestions, e.g., suggestions of body immobility, hallucination, and amnesia.) Furthermore, (*a*) some subjects who manifest these supposed characteristics of trance are unresponsive to test suggestions (Barber, 1957, 1963); (*b*) some subjects respond to test suggestions without manifesting these supposed signs of trance (Barber & Calverley, 1962, 1963a, 1963b, 1963c; Klopp, 1961); and (*c*) when subjects who are responsive to test suggestions show these trance characteristics, the characteristics can be removed by explicitly or tacitly suggesting to the subject that he no longer show them, and many subjects will continue to manifest a high level of response to test suggestions (Barber, 1958, 1962c; Gill & Brenman, 1959, p. 36; S. Fisher, 1954). These and other considerations presented elsewhere (Barber, 1964a) suggest that, at the present time, the term hypnotic state or trance is too nebulous and too difficult to denote clearly to be used to explain the enhanced suggestibility that is characteristically produced by hypnotic induction procedures and by task motivational instructions.

There are alternative formulations available to explain the heightened level of response to test suggestions that is typically produced by these two sets of variables. One alternative formulation posits that the two sets of variables tend to facilitate suggestibility in that one (task motivational instructions) includes explicit suggestions and the other (hypnotic induction) includes both explicit and implicit suggestions to the effect that the subject can manifest high response to test suggestions and is expected to cooperate and to try to respond maximally.

This alternative formulation is based on the following reasoning. Hypnotic induction procedures typically include explicit motivational suggestions together with suggestions that the subject can now easily respond to test suggestions (Variables 2 and 4 as listed in Table 1). However, even when such suggestions are not included in the hypnotic induction, the remaining components of the induction—defining the situation to the subject as hypnosis and administering suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, and sleep (Variables 1 and 3 as listed in Table 1)—may be sufficient to constitute an implicit task motivational procedure. It appears possible that when literate individuals in Western culture are told that they are participating in a hypnosis experiment and that they are going to be hypnotized, they are being told implicitly that they are participants in an unusual and important experiment in which they are expected to cooperate and, if they cooperate, they will find it easy to respond to test suggestions. Furthermore, since literate individuals are aware that suggestions of eye-heaviness, relaxation, drowsiness, and sleep are supposed to induce hypnosis, the administration of such suggestions serves to emphasize to them that the situation is hypnosis in which high response to test suggestions is not only possible but is also desired and expected.

A deduction from the above formulation can be tested experimentally. If Variables 1 and 3 (see Table 1) facilitate responsiveness to test suggestions by implicitly suggesting to the subject that he can respond well and is expected to try to respond maximally, then an experimental treatment consisting of these two variables should facilitate suggestibility only among subjects who are familiar with the traditional implications of the word hypnosis, that is, who associate the word hypnosis with high suggestibility and who believe that suggestions of eye-heaviness, relaxation, drowsiness, and sleep are effective in producing hypnosis. This deduction can be tested with a group of subjects, say nursery school or elementary school children, of whom some are and some are not familiar with the traditional connotations of the word hypnosis. The present formulation predicts that: (*a*) Both a task motivational treatment and a treatment consisting of Variables 1 and 3 will facilitate response to test suggestions among children who are acquainted with the cultural connotations of the word hypnosis. (*b*) A task motivational treatment, but *not* a treatment consisting of Variables 1 and 3, will facilitate suggestibility among children who have not as yet learned to associate the word hypnosis with high suggestibility and who have not as yet learned that suggestions of eye-heaviness, relaxation, drowsiness, and sleep are supposed to produce hypnosis and high suggestibility.

Effective Variables in "Hypnotic Induction" and "Task Motivational Instructions"

To state that both hypnotic induction procedures and task motivational instructions typically produce heightened response to test suggestions does not answer a crucial question: which of the many independent variables subsumed under these broad categories are instrumental and which extraneous in producing this effect? As Table 1 shows, the hypnotic induction procedures used in the experiments reviewed above typically included at least four independent variables. It may be that only one or two of these four variables are effective in facilitating suggestibility. Task motivational instructions also include more than one independent variable. The task motivational instructions used in the above experiments were typically comprised of: (*a*) motivational suggestions per se: "How well you do on the tests which I will give you depends entirely upon your willingness to try. . . . What I ask is your cooperation . . ."; and (*b*) suggestions that the subject can easily perform well and can easily experience the suggested effects: "Everyone passed these tests when they tried . . . you can easily imagine and do the interesting things I tell you. . . ." The first isolable variable included in task motivational instructions (motivational suggestions per se) is very similar if not identical to Variable 2 in hypnotic induction procedures, and the second variable included in task motivational instructions (suggestions that it is easy to respond well) is very similar if not identical to Variable 4 typically included in hypnotic inductions (see Table 1). Were both of these variables instrumental in facilitating suggestibility or was one variable effective and the other relatively ineffective?

A series of recent experiments, summarized below, were conducted (*a*) to ascertain which of the four independent variables typically included in hypnotic induction procedures are instrumental in facilitating response to test suggestions, and (*b*) to delineate the interactive effects of these variables in various combinations. As a by-product, these experiments also evaluated the effects on suggestibility of the two independent variables included in task motivational instructions that are practically indistinguishable from Variables 2 and 4 that are included in hypnotic inductions.

Effects of Variables 1 and 4. Glass and Barber (1961) presented data indicating that a higher level of response to the BSS is obtained when the experimental situation is defined to the subject as hypnosis rather than as "a control experiment involving a test of imagination." In a more extensive experiment, Barber and Calverley (1964k) appraised the main effects and the interactive effects on suggestibility of defining the situation to the subject as hypnosis (Variable 1) and also of suggesting to the subject that he will find it easy to perform well on the tests (Variable 4). The Barber and Calverley experiment was designed as a 2 × 2 factorial with each independent variable at two levels: definition of the experimental situation as hypnosis or as control, and suggestions that it would be easy or difficult to respond to the tests. Ninety-six students, who were required to participate in the experiment in fulfillment of course requirements, were randomly assigned to four experimental groups with 24 subjects per group. Groups A and B were told "You have been selected by chance to be in the hypnosis group and you are going to be hypnotized," whereas Groups C and D were told "You have been selected by chance to be in the control group and you will not be hypnotized." Groups A and C were next told that it was easy to respond to the tests whereas Groups B and D were told that it was difficult to pass the tests. Immediately following these instructions all subjects were assessed on the BSS (as summarized in Table 2). (To control possible experimenter bias in assessing suggestibility among the various groups, the BSS was administered to all subjects by means of a tape-recording of the experimenter's voice.) The main effects of both independent variables were significant: higher objective and subjective scores on the BSS were obtained when the situation was defined as hypnosis rather than as control, and when response to the tests was described as easy rather than as difficult. The suggestibility-enhancing effects of the independent variables were additive: the level of suggestibility was highest when the situation was defined to the subject as hypnosis and response to the tests was described as easy (Group A); next highest when *either* the situation was defined as hypnosis *or* response to the tests was described as easy (Groups B and C); and lowest when the situation was defined to the subject as a "control experiment" and response to the tests was described as difficult (Group D).

Barber and Calverley noted that further studies are required to specify the intervening processes by which describing the experiment as hypnosis or stating that it is easy to respond to test suggestions produce a higher level of suggestibility than describing the situation as a control experiment or stating that it is difficult to respond to the tests. They suggested that such studies should attempt to assess the subject's interpretation of the instructions, and they postulated that, (*a*) when subjects are told that they have been assigned to the control group or that the test suggestions are difficult to pass, they may interpret these statements to mean that they are *not* expected to respond to suggestions of arm levitation, body immobility, selective amnesia, and so on, whereas (*b*) when they are told that they have been assigned to the hypnosis group or that the test suggestions are easy to pass they may interpret these statements to mean that they are expected to respond positively to such suggestions.

Effects of Variables 1, 2, and 4. Barber and Calverley (1965a) subsequently conducted an experiment with 136 subjects which replicated and extended the above study. This experiment was designed as a 2 × 2 × 2 factorial with three independent variables—Variables 1, 2, and 4 as listed in Table 1—each at two levels. Instructions for the two levels of Variable 1 were "You have been selected by chance to be in the hypnosis group and you are going to be hypnotized" versus "You have been selected by chance to be in the control group . . . you will not be hypnotized . . . . I'm going to give you several tests of imagination." The two levels of Variable 2 consisted of the presence and absence of motivational instructions: "Please try hard. . . . If you don't try, the experiment will be worthless. . . ." Instructions for one level of Variable 4 (suggestions that it is easy to respond to the tests) were the same as in the preceding experiment; however, instructions for the second level of Variable 4 differed from that of the preceding experiment in that, instead of suggestions that it would be difficult to respond to the tests, nothing was said concerning the ease or difficulty of responding. All subjects were assessed on subjective and subjective responses to the BSS (see Table 2). (To control experimenter bias, the BSS was administered to all subjects by means of a tape recording.)

Confirming the results of the preceding experiment, subjects told "You are in the hypnosis group" obtained higher objective and subjective scores on the BSS than those told "You are in the control group. . . . I'm going to give you several tests of imagination." The main effect of Variable 2 (motivational instructions) was significant. However, a significant triple interaction indicated that motivational instructions consistently facilitated response to the BSS when the situation was defined as control but did not consistently facilitate response when the situation was defined as hypnosis. The main effect of Variable 4 (suggesting or not suggesting that it is easy to respond to the tests) was nonsignificant. However, the significant triple interaction indicated that suggestions that it is easy to respond to the tests tended to enhance the scores on the BSS when the situation was defined as a "control experiment" and motivational instructions were also administered. Multiple comparisons of the group means showed that correspondingly high objective and subjective scores on the BSS can be produced either by (*a*) defining the situation to the subject as hypnosis (Variable 1), with or without Variables 2 and 4, or by (*b*) defining the situation as a "control experiment involving a test of imagination," and in addition administering motivational instructions and suggestions that it is easy to respond to the tests (Variables 2 and 4).

Effects of Variables 1, 3, and 2 and 4. In addition to Variables 1, 2, and 4 which were evaluated in the above experiment, hypnotic induction procedures typically include an additional variable, namely, repeated suggestions of relaxation, drowsiness, and sleep (Variable 3). Barber and Calverley (1965a) subsequently conducted a further study to ascertain whether this additional variable is effective in elevating responsiveness to test suggestions as compared to (*a*) simply defining the experimental situation to the subject as hypnosis (Variable 1), and to (*b*) administering motivational instructions together with suggestions that the subject can easily respond to the tests (Variables 2 and 4). Seventy-four college student volunteers were randomly assigned to five experimental groups. With respect to Groups A, B, C, and D the situation was defined as hypnosis and suggestions of relaxation, drowsiness, and sleep were administered for 0, 1, 5, and 10 minutes, respectively. With respect to Group E, the situation was defined as a control experiment, response to the tests was described as easy, and motivational instructions were administered. Objective and subjective responses to the BSS constituted the dependent variables (see Table 2). (To control possible experimenter bias in administering the test suggestions, the BSS was given to all subjects by a tape recording.) The mean objective and subjective scores of Group A were significantly smaller than the mean scores of Groups B, C, D, and E and the mean scores of the latter four groups did not differ significantly from each other. The results thus indicated that: (*a*) Variable 3 (suggestions of relaxation, drowsiness, and sleep administered for either 1, 5, or 10 minutes) in combination with Variable 1 (defining the situation as hypnosis) is more effective than Variable 1 alone in producing high response to test suggestions; (*b*) suggestions of relaxation, drowsiness, and sleep administered for 5 minutes or for 10 minutes are not noticeably more effective in facilitating response to the BSS than the same suggestions administered for 1 minute; and (*c*) Variables 2 and 4 in combination (motivational instructions together with suggestions that the subject will find it easy to perform well on the tests) appear to be as effective as Variables 1 and 3 in combination in facilitating response to test suggestions.

Effects of Variables 3, 2 and 4. In a further experiment, Barber and Calverley (in press) held Variable 1 constant (that is, the situation was defined to all subjects as hypnosis) while experimentally manipulating Variables

3, 2, and 4. Fifty-six student nurse volunteers were randomly assigned to four experimental groups with 14 per group. The experimental treatment for Groups A and B included Variable 2 (motivational instructions) and Variable 4 (suggestions that it is easy to respond to the tests). The experimental treatment for Groups A and C included Variable 3 (5 minutes of suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, and sleep). The dependent variables were responses to the BSS. (To control possible experimenter bias, the BSS was administered by a tape recording.) As compared to the control treatment that did not include either Variables 3 or 2 and 4 (Group D), the treatment comprised of Variables 2 and 4 (Group B) and also the treatment comprised of Variable 3 (Group C) were both effective in augmenting objective and subjective scores on the BSS. However, the suggestibility-enhancing effects of Variable 3 and of Variables 2 and 4 were not additive: Group A (Variables 3, 2, and 4) did not show a higher level of response than Group C (Variable 3) or Group B (Variables 2 and 4). Stated otherwise, (*a*) motivational instructions together with suggestions that it is easy to respond to the tests and (*b*) suggestions of relaxation, drowsiness, and sleep were equally effective in producing a high level of response to test suggestions, and *a* and *b* in combination were not more effective than either *a* or *b* alone.

Conclusions from the above experiments. In general, the results of the four experiments summarized in this section appear to indicate the following:

1. It appears that simply defining the experimental situation to the subject as hypnosis tends to produce an increment in responsiveness to test suggestions. With other experimental variables counterbalanced across groups, subjects told that they are to be hypnotized tend to be more responsive to suggestions of arm levitation, thirst hallucination, body immobility, selective amnesia, and so on, than subjects told that they are to receive a test of imagination and will not be hypnotized.

2. Further increments in response to test suggestions can be produced by administering the following suggestions or instructions, in addition to defining the situation to the subject as hypnosis: suggestions of eye-heaviness, relaxation, drowsiness, and sleep; or motivational instructions together with suggestions that it is easy to respond to the tests.

3. When the situation is defined to the subject as hypnosis, motivational instructions together with suggestions that it is easy to respond to the tests (Variables 2 and 4) appear to be as effective in facilitating response to test suggestions as suggestions of eye-heaviness, relaxation, drowsiness, and sleep (Variable 3). The enhanced suggestibility produced by these two sets of variables, however, does not appear to be additive; an experimental treatment that includes Variable 3 and also Variables 2 and 4 does not seem to produce a higher level of response to the BSS than an experimental treatment that includes only Variables 3 or only Variables 2 and 4.

4. Although defining the situation to the subject as hypnosis tends to produce a higher level of response to test suggestions than defining the situation as a control experiment, this does not mean that a high level of response cannot be elicited when the situation is defined as control. However, to elicit high response to test suggestions from subjects who have been told that they have been assigned to a control group and will not be hypnotized, it appears necessary to administer additional instructions and suggestions, for instance, motivational instructions and suggestions that it is easy to respond to the tests. Furthermore, the available data appear to indicate that comparable high levels of response to test suggestions can be elicited either by (*a*) administering suggestions of relaxation, drowsiness, and sleep in a situation defined to the subject as hypnosis, or by (*b*) administering suggestions that the tests are easy to pass together with motivational instructions in a situation defined to the subject as a "control experiment involving a test of imagination."

It needs to be emphasized that these conclusions are derived from experiments that employed response to the BSS as the dependent variable. Further studies are needed to confirm and extend these findings using as dependent variables not only response to the BSS but also response to other test suggestions not included in this scale such as suggestions of analgesia, age-regression, and visual and auditory hallucinations.

PART II. EFFECTS ON SUGGESTIBILITY OF PERSONALITY VARIABLES AND SEVERAL RELATIVELY SUBTLE INSTRUCTIONAL-SITUATIONAL VARIABLES

To formulate a general theory of hypnotic behavior it is necessary to specify which of the many independent variables typically present in hypnotic experimental situations play an important role and which play a relatively minor or unimportant role in determining response to test suggestions of body immobility, hallucination, amnesia, and so on. The experiments reviewed above point to several variables that seem to exert an important effect. Two series of experiments are reviewed next. The first series succeeded in isolating several additional instructional-situational variables that also seem to affect response to test suggestions. The second series indicates that personality variables and several instructional-situational variables, which have been postulated to play a significant role in determining response to test suggestions, do not seem to exert a major effect when tested experimentally.

Variables Affecting Response

Experimenter's tone of voice in administering suggestions. Barber and Calverley (1964d) assessed response to test suggestions while varying the tone of voice in which the suggestions were presented. Student nurse volunteers were randomly assigned to Groups A and B with 41 to each group, and the experimenter presented the BSS in a forceful tone to Group A and in a lackadaisical tone to Group B. (Two psychologists and one sociologist, who were not informed as to the design of the experiment, agreed in rating one presentation as firm and forceful and the other presentation as listless and lackadaisical.) Group A on the average passed approximately 4 of the 8 test suggestions whereas Group B passed approximately 2. Further, although the subjects were randomly allocated from a homogeneous population to the two experimental groups, the proportion manifesting "high" suggestibility (scores of 5.5 or above on the 8-point BSS) varied from 34% (Group A) to 12% (Group B); and the proportion manifesting "low" suggestibility (scores of 1.5 or less) varied from 20% (Group A) to 49% (Group B).

Positive and negative attitudinal-motivational instructions. In a further experiment Barber and Calverley (1964e) obtained data indicating that high and low levels of response to test suggestions can be obtained by administering instructions designed to produce positive and negative attitudes and motives with respect to the test situation. Twenty-four student nurses, who had been rated on response to the BSS in a preliminary session, were randomly allocated to three experimental groups with 8 per group. Group 1 (Positive instructions) was retested on the BSS after receiving motivational instructions together with suggestions that it is easy to respond to the tests. Group 2 (Neutral instructions) was retested on the BSS in the same way as all subjects had been tested in the preliminary session, that is, without special instructions. Group 3 (Negative instructions) was retested on the BSS after being told the following by a prestigeful person (the Supervisor of Student Nurses):

> It's being rumored by doctors and administrators, and I don't know who else, that nursing students are too easily directed and easily led in their responses to Dr. Calverley's suggestions. It's kind of shocking and discouraging to hear that the students, up to this time, are so easily directed and can't decide things for themselves. We've got a job to do; to impress the administrators and doctors around here with the fact that nursing students are not as gullible and as easily directed as they appear to have been showing during this research study. Well it sure is up to each of you as to how easily led people around here think student nurses are.

Although subjects allocated to Groups 1, 2, and 3 had obtained very similar mean scores on the BSS in the preliminary session (passing approximately 3 of the 8 test suggestions), their scores on retest differed markedly. On the average, Group 1 (Positive instructions) passed 4 of the test suggestions, Group 2 (Neutral instructions) passed 2, and with the exception of one subject who passed one test, all subjects in Group 3 (Negative instructions) obtained zero scores on the BSS. It should be noted that, depending on whether positive or negative instructions were administered, the number of subjects who were *not* suggestible in the formal experimental situation (that is, who responded to none of the eight test suggestions included in the BSS) varied from 88% (Group 3) to 0% (Group 1).

What subjects are told is the purpose of the experiment. Barber and Calverley (1964c) demonstrated that clear-cut variations in response to standardized test suggestions can be produced by varying what the subject is told is the purpose of the experiment. Seventeen student nurses were randomly assigned to Group A and 16 to Group B. Group A was told that the purpose of the experiment was to test imagination whereas Group B was told that the purpose was to test gullibility. Both groups were assessed on response to the BSS. (To control possible experimenter bias in administering the test suggestions, the BSS was administered to all subjects by means of a tape recording.) Forty-one percent of the subjects in Group A and only 6% in Group B obtained relatively high scores of 5 or above on the 8-point BSS.

Suggestions implying that the procedure to be used is or is not effective in producing "hypnosis." Barber and Calverley (in press) recently reported an experiment that included two independent variables, each at two levels. The two levels of one independent variable were suggestions implying that the procedure to be employed is or is not effective in producing hypnosis. The two levels of the second independent variable were presence versus absence of suggestions of eye-heaviness, relaxation, drowsiness, and sleep administered for 5 minutes by a tape recording. Eighty-four student nurses and practical nurses were required to participate in the hypnosis experiment as part of course requirements. These "captive" subjects were randomly assigned to the four cells comprising the 2 × 2 factorial, with 21 per cell. The dependent variables were objective and subjective scores on the BSS (Table 2). (To control possible experimenter bias, the BSS was administered to all subjects by means of a tape recording.) The overall results were: (*a*) Suggestions implying that the procedure to be used is effective in producing hypnosis gave rise to a higher level of objective and subjective responses to the BSS than suggestions implying that the procedure to be used is not effective in inducing hypnosis. (b) In harmony with previous studies, suggestions of eye-heaviness, relaxation, drowsiness, and sleep were effective in elevating response to test suggestions. (*c*) Suggestions implying that the procedure to be used is or is not effective in inducing hypnosis exerted a more potent effect on response to the BSS than the presence or absence of suggestions of eye-heaviness, relaxation, drowsiness, and sleep.

Variables Not Exerting an Important Effect on Response

Transsituational personality characteristics. A large number of investigations have been conducted to test the hypothesis that enduring characteristics of personality exert an important effect on hypnotizability or suggestibility. In these studies hypnotizability and suggestibility were generally denoted operationally by assessing subjects' level of response to standardized test suggestions administered with and without a preceding hypnotic induction procedure, respectively. Personality variables were assessed by: (*a*) projective techniques (Doland, 1953; Sarbin & Madow, 1942; Stukat, 1958); (*b*) ratings, autobiographies, and clinical interviews (Gill & Brenman, 1959, pp. 81–83; Hilgard & Hilgard, 1962); (*c*) self-report questionnaires concerning the nature and degree of the subject's previous imaginative-fantasy or hypnotic-like experiences (As, 1962; Barber & Calverley, 1965b; Shor, Orne, & O'Connell, 1962); and (*d*) standardized inventories such as the MMPI, the Guilford-Zimmerman Temperament Survey, the Maudsley Personality Inventory, the Edwards Personal Preference Schedule, and the Leary Interpersonal Check List (Barber, 1956; Barber & Calverley, 1964g, 1964h, 1964i; Faw & Wilcox, 1958; Furneaux & Gibson, 1961; Hilgard & Bentler, 1963; Secter, 1961; Weitzenhoffer & Weitzenhoffer, 1958). The many investigations in this area, including the ones cited and approximately 50 others which were recently reviewed elsewhere (Barber, 1964d), either failed to find significant differences in personality characteristics between subjects ranking high and low on hypnotizability or suggestibility or else reported conflicting findings.

To what can we attribute this overall failure to find attributes of personality which are consistently related to responsiveness to test suggestions? It may be that subtle aspects of personality that are related to hypnotizability or suggestibility are not tapped by present-day methods of personality assessment. However, it could also be that response to test suggestions is primarily situationally determined and that transsituational characteristics of personality may play only a very small role or may play a role only in interaction with situational variables. The latter hypothesis is in line with a series of experiments, reviewed earlier in the present paper, which indicated that variations in the experimental situation—e.g., variations in the experimenter's tone of voice when administering the test suggestions and variations in the words used to describe the purpose of the experiment to the subjects—produce clear-cut variations in subjects' responsiveness to test suggestions.

Eyes closed or open. In hypnotic experiments subjects are almost always assessed on response to test suggestions with their eyes closed. Implicit in this method of assessment appears to be the assumption that subjects are more suggestible when their eyes are closed rather than open. Barber and Calverley (1965a) evaluated this assumption as follows. Twenty-four subjects were assessed on the BSS without receiving special instructions; 12 of these, selected at random, were assessed with eyes open and the others with eyes closed. Two additional groups of 12 subjects were rated on the BSS with eyes closed and open, respectively, after receiving a standardized hypnotic induction procedure (as outlined in Table 1). (To control variations in the experimenter's tone of voice when administering the test suggestions, the BSS was given to all subjects by means of a tape recording.) Under both the hypnotic induction and the nonhypnotic induction conditions subjects were slightly more responsive but not significantly more responsive to the test suggestions when their eyes were open rather than closed. Further studies are needed to ascertain whether this variable (eyes closed versus eyes open) exerts an important effect on response to test suggestions other than those included in the BSS, e.g., suggestions of age-regression or suggestions of enhanced memory for previous events.

Personal versus relatively impersonal administration of suggestions. In hypnosis experiments the hypnotic induction procedure and the test suggestions are almost always administered orally. Implicit in this method of administration is the assumption that a higher level of suggestibility is elicited when the test suggestions are presented in a more personal manner (as spoken by the experimenter) rather than in a relatively impersonal manner (for instance, by a tape recording). Barber and Calverley (1964b) evaluated this assumption in two experiments. In the first experiment 84 student nurses were assigned at random to two groups of 42 subjects and both groups were assessed on response to the BSS. To one group the test suggestions were presented orally by the experimenter and to the other group the identical test suggestions were presented by means of a tape recording of the experimenter's voice. The mean objective and subjective scores were slightly higher but not significantly higher under the recorded rather than under the spoken presentation. In the second experiment, 66 male dental students were randomly assigned to two groups of 33 subjects. A standardized hypnotic induction procedure and the BSS were administered to one group orally and to the other by a tape recording. The mean objective scores under recorded and spoken presentation were identical and the mean subjective scores were practically identical (cf., Hoskovec, Svorad, & Lanc, 1963).

OVERVIEW

The experiments reviewed in this paper point to several variables that appear to play an important role in producing positive responses to suggestions of hallucinations, analgesia, amnesia, body immobility, and so on, that is, in producing behaviors of the type traditionally labeled as hypnotic. Broadly speaking, it appears that an experimental treatment of the type labeled as a hypnotic induction and also an experimental treatment labeled as task motivational instructions are both effective in eliciting hypnotic behaviors. However, both of these experimental treatments include several independent variables. Task motivational instructions include motivational instructions per se plus suggestions that it is easy to respond to assigned tasks. Hypnotic induction procedures typically include the two variables included in task motivational instructions plus two additional variables: definition of the situation to the subject as hypnosis; and suggestions of eye-heaviness, relaxation, drowsiness, and sleep.

A series of experiments has been conducted to ascertain the relative importance of these four variables in facilitating response to test suggestions. These experiments appear to indicate that:

1. Defining the situation to the subject as hypnosis tends to produce a higher level of suggestibility than defining it as a control experiment.

2. When the situation is defined to the subject as hypnosis, correspondingly high levels of response to test suggestions can be evoked either by (*a*) administering motivational instructions together with suggestions that it is easy to respond to the tests, (*b*) administering suggestions of eye-heaviness, eye-closure, relaxation, drowsiness, and sleep, or (*c*) administering an experimental treatment that includes both *a* and *b*.

3. It is also possible to produce a high level of suggestibility when the situation is defined to the subject as a control experiment, provided that motivational instructions and suggestions that it is easy to respond to the tests are also administered.

4. It appears that equally high levels of response to test suggestions can be elicited either by (*a*) administering suggestions of eye-heaviness, relaxation, drowsiness, and sleep to subjects who have been told that they are to be hypnotized, or by (*b*) administering motivational instructions together with suggestions that it is easy to perform well on the tasks to subjects who have been told that they are not to be hypnotized but are to serve in a control group and are to receive tests of imagination.

The experiments cited also point to several additional instructional-situational variables that influence response to test suggestions. These include: (*a*) the tone of voice in which the test suggestions are presented; (*b*) whether prospective subjects are told, by a prestigeful person, that they should or should not respond to the suggestions that they will receive in the experiment; (*c*) what subjects are told is the purpose of the experiment; and (*d*) whether subjects, who have been told that they are participants in a hypnosis experiment, are given implicit suggestions that the procedure to be used is or is not effective in inducing hypnosis. The above experiments also indicate that the following variables, which have been postulated to play an important role in determining response to test suggestions, do not seem to be important when tested experimentally: (*a*) the trans-situational personality characteristics of the subject; (*b*) whether the test suggestions are administered to the subject with his eyes closed or open; and (*c*) whether the test suggestions are administered in a relatively impersonal manner (by a tape recording) or in a more personal manner (orally by the experimenter).

LIMITATIONS OF THE INVESTIGATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The results obtained in the above investigations may be valid only under certain limiting conditions. For instance, informing the subject that he is to be hypnotized may facilitate suggestibility only in certain classes of subjects, such as the college students and nursing students that were used in almost all of the experiments reviewed above, and may not facilitate suggestibility in other types of subjects, for instance, working-class adults, aged individuals or hospitalized mental patients (cf., Barber, Karacan, & Calverley, 1964). This consideration leads to the following general criticism of the investigations cited.

In most of the experiments volunteer subjects who had not been pretested on suggestibility were assigned at random to the treatments. In other experiments, however, subjects were assigned randomly to treatments from a population that had previously manifested high response to the BSS. Further, although most of the experiments were conducted with volunteers, some were conducted with captive subjects. In addition, subjects in some experiments were college undergraduates whereas in other experiments the subjects were nursing students. The results obtained with previously untested subjects may or may not apply to subjects who have previously shown high response to test suggestions; the results obtained with volunteers may or may not apply to captives (cf., Boucher & Hilgard, 1962); the results obtained with college students may or may not apply to nursing students. Further studies are needed to ascertain whether results obtained with one type of adult sample can be generalized to other types of adult samples, and whether results which are found to hold for various samples of adults also apply to children, to adolescents, and to the aged (cf., Barber & Calverley, 1963a).

A second general criticism of the above investigations is that no attempt was made to ascertain what motives led subjects to volunteer for the experiment, how subjects perceived the experimental situation, and how they interpreted the instructions they were given. For instance, how did the subjects interpret instructions stating that they were to be hypnotized or instructions stating that they had been assigned to a control group and would not be hypnotized but would receive a test of imagination? Further studies are needed which carefully attempt to assess the motives which lead subjects to volunteer and also to assess how subjects perceive the experimental situation, how they interpret the instructions they receive, and how these "internal" variables affect objective response.

Rigorous research is also needed to appraise the effects of several additional subject variables. For instance, to what extent is response to test suggestions affected by subjects' attitude toward hypnosis? A series of experimental studies (Barber & Calverley, 1963c; Melei & Hilgard, 1964; Rosenhan & Tomkins, 1964; Secter, 1960; White, 1937; Willey, 1951) indicates that subjects holding positive attitudes toward hypnosis tend to score slightly higher, but not always significantly higher, on test suggestions than subjects holding negative attitudes. However, these studies are not conclusive and further experimentation is needed to clarify the effects of this variable. Additional research is also needed to delineate the effects on suggestibility of the events occurring during subject recruitment and the events occurring when the subject first enters the experimental room, e.g., the subject's initial reactions to the experimental setting and to the person of the experimenter.

The subject variables mentioned above should be considered in relation to a number of experimenter variables. For instance, what are the effects of shared beliefs on the part of subject and experimenter concerning the "power of suggestions" or the "power of hypnosis"? What variance is introduced into the subject's responsiveness by the person of the experimenter—his personality characteristics, his "emotional warmth," his ethnicity, age, and sex, and his attitudes toward the experiment, toward the subject, and toward his own role as an experimenter? What bias is introduced into the experimental outcome by the experimenter's expectations, his investment in a given experimental outcome, his facial expressions, his posture and movements, and the quality, pitch, and resonance of his voice? To what extent are the experimental instructions confounded with the characteristics of the experimenter administering the instructions?

Additional research is also needed to delineate further the gross variables that were evaluated in the experiments described above. For instance, Variable 3, as evaluated in the above experiments, typically included (*a*) suggestions of eye-heaviness and eye-closure, (*b*) suggestions of relaxation, and (*c*) suggestions of drowsiness and sleep. It may be that only one of these three isolable factors included in Variable 3 was effective in facilitating response to test suggestions and that the other two factors did not exert an important effect. Research is needed not only to specify further the factors involved in each of the gross variables assessed in the above experiments but also (*a*) to isolate additional variables that affect response to test suggestions, (*b*) to assign weights to the effective variables and to their interactions, and (*c*) to specify the processes by which the antecedent variables are instrumental in producing behaviors of the type which have been traditionally viewed as due to or associated with hypnosis.

Finally, although there is a need for a formal theory of hypnotic behavior, the writer sees no grounds for believing that an adequate theory can arise from anything but the patient analyses of experimentally manipulated variables. There is much to be done in straightforward empirical research, following up leads that are now available, before it will be profitable to attempt a formal theory. One can venture to predict that, when sufficient experimental data have accumulated, any simple formula to explain hypnotic behavior will be open to serious question. The influence of any one variable will be found to depend upon the context of other variables. The determinants of response to test suggestions will be found to be multiple and complex. Viewed in this perspective the present state of research seems primitive indeed. We have just begun to break the ground.

REFERENCES

As, A. Non-hypnotic experiences related to hypnotizability in male and female college students. *Scandinavian Journal of Psychology,* 1962, **3,** 112–121.

Barber, T. X. A note on "hypnotizability" and personality traits. *Journal of Clinical and Experimental Hypnosis,* 1956, **4,** 109–114.

Barber, T. X. Hypnosis as perceptual-cognitive restructuring: III. From somnambulism to autohypnosis. *Journal of Psychology,* 1957, **44,** 299–304.

Barber, T. X. Hypnosis as perceptual-cognitive restructuring: II. "Post"-hypnotic behavior. *Journal of Clinical and Experimental Hypnosis,* 1958, **6,** 10–20.

Barber, T. X. Antisocial and criminal acts induced by "hypnosis": A review of clinical and experimental findings. *Archives of General Psychiatry,* 1961, **5,** 301–312. (a)

Barber, T. X. Physiological effects of "hypnosis." *Psychological Bulletin,* 1961, **58,** 390–419. (b)

Barber, T. X. Experimental controls and the phenomena of "hypnosis": A critique of hypnotic research methodology. *Journal of Nervous and Mental Disease,* 1962, **134,** 493–505. (a)

Barber, T. X. Hypnotic age-regression: A critical review. *Psychosomatic Medicine,* 1962, **24,** 286–299. (b)

Barber, T. X. Toward a theory of hypnosis: Posthypnotic behavior. *Archives of General Psychiatry,* 1962, **7,** 321–342. (c)

Barber, T. X. Toward a theory of "hypnotic" behavior: The "hypnotically induced dream." *Journal of Nervous and Mental Disease,* 1962, **135,** 206–221. (d)

Barber, T. X. The effects of "hypnosis" on pain: A critical review of experimental and clinical findings. *Psychosomatic Medicine,* 1963, **25,** 303–333.

Barber, T. X. "Hypnosis" as a causal variable in present-day psychology: A critical analysis. *Psychological Reports,* 1964, **14,** 839–842. (a)

Barber, T. X. Hypnotically hallucinated colors and their negative after-images. *American Journal of Psychology,* 1964, **77,** 313–318. (b)

Barber, T. X. Hypnotic "colorblindness," "blindness," and "deafness": A review of research findings. *Diseases of the Nervous System,* 1964, **25,** 529–537. (c)

Barber, T. X. Hypnotizability, suggestibility, and personality: V. A critical review of research findings. *Psychological Reports,* 1964, **14,** 299–320. (d)

Barber, T. X. Physiological indices of "hypnosis": A critical evaluation. Harding, Mass.: Medfield Foundation, 1964. (Mimeo) (e)

Barber, T. X. Toward a theory of "hypnotic" behavior: Positive visual and auditory hallucinations. *Psychological Record,* 1964, **14,** 197–210. (f)

Barber, T. X. Effects of "hypnosis" on learning and recall: A methodological critique. *Journal of Clinical Psychology,* 1965, **21,** 19–25. (a)

Barber, T. X. Physiological effects of "hypnotic suggestions": A critical review of recent research (1960–1964). *Psychological Bulletin,* 1965, **63,** 201–222. (b)

Barber, T. X. Effects of "hypnosis" and motivational suggestions on strength and endurance: A critical review of research studies. *British Journal of Social and Clinical Psychology,* in press.

Barber, T. X., & Calverley, D. S. "Hypnotic behavior" as a function of task motivation. *Journal of Psychology,* 1962, **54,** 363–389.

Barber, T. X., & Calverley, D. S. "Hypnotic-like" suggestibility in children and adults. *Journal of Abnormal and Social Psychology,* 1963, **66,** 589–597. (a)

Barber, T. X., & Calverley, D. S. The relative effectiveness of task motivating instructions and trance induction procedure in the production of "hypnotic-like" behaviors. *Journal of Nervous and Mental Disease,* 1963, **137,** 107–116. (b)

Barber, T. X., & Calverley, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of task motivating instructions and attitudes toward hypnosis. *Journal of Abnormal and Social Psychology,* 1963, **67,** 557–565. (c)

Barber, T. X., & Calverley, D. S. An experimental study of "hypnotic" (auditory and visual) hallucinations. *Journal of Abnormal and Social Psychology,* 1964, **68,** 13–20. (a)

BARBER, T. X., & CALVERLEY, D. S. Comparative effects on "hypnotic-like" suggestibility of recorded and spoken suggestions. *Journal of Consulting Psychology*, 1964, **28**, 384. (b)

BARBER, T. X., & CALVERLEY, D. S. Definition of the situation as a variable affecting "hypnotic-like" suggestibility. *Journal of Clinical Psychology*, 1964, **20**, 438–440. (c)

BARBER, T. X., & CALVERLEY, D. S. Effect of *E*'s tone of voice on "hypnotic-like" suggestibility. *Psychological Reports*, 1964, **15**, 139–144. (d)

BARBER, T. X., & CALVERLEY, D. S. Empirical evidence for a theory of "hypnotic" behavior: Effects of pretest instructions on response to primary suggestions. *Psychological Record*, 1964, **14**, 457–467. (e)

BARBER, T. X., & CALVERLEY, D. S. Experimental studies in "hypnotic" behaviour: Suggested deafness evaluated by delayed auditory feedback. *British Journal of Psychology*, 1964, **55**, 439–446. (f)

BARBER, T. X., & CALVERLEY, D. S. Hypnotizability, suggestibility, and personality: I. Two studies with the Edwards Personal Preference Schedule, the Jourard Self-Disclosure Scale, and the Marlow-Crowne Social Desirability Scale. *Journal of Psychology*, 1964, **58**, 215–222. (g)

BARBER, T. X., & CALVERLEY, D. S. Hypnotizability, suggestibility, and personality: III. A study using teachers' ratings of children's characteristics. *Journal of Psychology*, 1964, **57**, 275–280. (h)

BARBER, T. X., & CALVERLEY, D. S. Hypnotizability, suggestibility, and personality: IV. A study with the Leary Interpersonal Check List. *British Journal of Social and Clinical Psychology*, 1964, **3**, 149–150. (i)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of "hypnotic" behavior: An experimental study of "hypnotic time distortion." *Archives of General Psychiatry*, 1964, **10**, 209–216. (j)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. *Journal of Abnormal and Social Psychology*, 1964, **68**, 585–592. (k)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of "hypnotic" behaviour: Enhancement of strength and endurance. *Canadian Journal of Psychology*, 1964, **18**, 156–167. (l)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of "hypnotic" behavior: Experimental analyses of suggested amnesia. Harding, Mass.: Medfield Foundation, 1964. (Mimeo) (m)

BARBER, T. X., & CALVERLEY, D. S. Empirical evidence for a theory of "hypnotic" behavior: Effects on suggestibility of five variables typically included in "hypnotic induction" procedures. *Journal of Consulting Psychology*, 1965, **29**, 98–107. (a)

BARBER, T. X., & CALVERLEY, D. S. Hypnotizability, suggestibility, and personality: II. Assessment of prior imaginative-fantasy experiences by the As, Barber-Glass, and Shor questionnaires. *Journal of Clinical Psychology*, 1965, **21**, 57–58. (b)

BARBER, T. X., & CALVERLEY, D. S. Empirical evidence for a theory of "hypnotic" behavior: The suggestibility-enhancing effects of motivational suggestions, relaxation-sleep suggestions, and suggestions that the subject will be effectively "hypnotized." *Journal of Personality*, in press.

BARBER, T. X., CHAUNCEY, H. H., & WINER, R. A. Effect of hypnotic and nonhypnotic suggestions on parotid gland response to gustatory stimuli. *Psychosomatic Medicine*, 1964, **26**, 374–380.

BARBER, T. X., & DEELEY, D. S. Experimental evidence for a theory of hypnotic behavior: I. "Hypnotic color-blindedness" without "hypnosis." *International Journal of Clinical and Experimental Hypnosis*, 1961, **9**, 79–86.

BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *Journal of Abnormal and Social Psychology*, 1962, **64**, 222–228.

BARBER, T. X., & HAHN, K. W., JR. Physiological and subjective responses to pain producing stimulation under hypnotically-suggested and waking-imagined "analgesia." *Journal of Abnormal and Social Psychology*, 1962, **65**, 411–418.

BARBER, T. X., & HAHN, K. W., JR. Experimental studies in "hypnotic" behavior: Physiological and subjective effects of imagined pain. *Journal of Nervous and Mental Disease*, 1964, **139**, 416–425.

BARBER, T. X., KARACAN, I., & CALVERLEY, D. S. "Hypnotizability" and suggestibility in chronic schizophrenics. *Archives of General Psychiatry*, 1964, **11**, 439–451.

BOUCHER, R. G., & HILGARD, E. R. Volunteer bias in hypnotic experimentation. *American Journal of Clinical Hypnosis*, 1962, **5**, 49–51.

BOWLES, J. W., JR., & PRONKO, N. H. Reversibility of stimulus function under hypnosis. *Journal of Psychology*, 1949, **27**, 41–48.

BRAVIN, M. Role-play and direct suggestion in hypnotically induced color blindness. Unpublished doctoral dissertation, University of Denver, 1959.

BROWN, R. R., & VOGEL, V. H. Psychophysiological reactions following painful stimuli under hypnotic analgesia contrasted with gas anesthesia and novocain block. *Journal of Applied Psychology*, 1938, **22**, 408–420.

DOLAND, D. J. A study of the relationship between primary suggestibility and certain measures of ego strength. Unpublished doctoral dissertation, Syracuse University, 1953.

ERICKSON, M. H. A study of clinical and experimental findings on hypnotic deafness: I. Clinical experimentation and findings. *Journal of General Psychology*, 1938, **19**, 127–150. (a)

ERICKSON, M. H. A study of clinical and experimental findings on hypnotic deafness: II. Experimental findings with a conditioned response technique. *Journal of General Psychology*, 1938, **19**, 151–167. (b)

ERICKSON, M. H., HERSHMAN, S., & SECTER, I. I. *The practical applications of medical and dental hypnosis*. New York: Julian Press, 1961.

FAW, V., & WILCOX, W. W. Personality characteristics of susceptible and unsusceptible hypnotic subjects.

Journal of Clinical and Experimental Hypnosis, 1958, **6**, 83–94.

FISHER, C. Studies on the nature of suggestion: I. Experimental induction of dreams by direct suggestion. *Journal of the American Psychoanalytic Association,* 1953, **1**, 222–255.

FISHER, S. The role of expectancy in the performance of posthypnotic behavior. *Journal of Abnormal and Social Psychology,* 1954, **49**, 503–507.

FOWLER, W. L. Hypnosis and learning. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 223–232.

FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *Journal of Abnormal and Social Psychology,* 1938, **33**, 453–475.

FURNEAUX, W. D., & GIBSON, H. B. The Maudsley Personality Inventory as a predictor of susceptibility to hypnosis. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 167–176.

GILL, M. M., & BRENMAN, MARGARET. *Hypnosis and related states.* New York: International Universities Press, 1959.

GLASS, L. B., & BARBER, T. X. A note on hypnotic behavior, the definition of the situation and the placebo effect. *Journal of Nervous and Mental Disease,* 1961, **132**, 539–541.

GORDON, J. E. (Ed.) *Handbook of hypnosis.* New York: Macmillan, in press.

HADFIELD, J. A. *The psychology of power.* London: Macmillan, 1924.

HAMMER, E. F. Post-hypnotic suggestion and test performance. *Journal of Clinical and Experimental Hypnosis,* 1954, **2**, 178–185.

HILGARD, E. R. Hypnosis. *Annual Review of Psychology,* 1965, **16**, in press.

HILGARD, E. R., & BENTLER, P. M. Predicting hypnotizability from the Maudsley Personality Inventory. *British Journal of Psychology,* 1963, **54**, 63–69.

HILGARD, JOSEPHINE R., & HILGARD, E. R. Developmental-interactive aspects of hypnosis: Some illustrative cases. *Genetic Psychology Monographs,* 1962, **66**, 143–178.

HOSKOVEC, J., SVORAD, D., & LANC, O. The comparative effectiveness of spoken and tape-recorded suggestions of body sway. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11**, 163–166.

IKAI, M., & STEINHAUS, A. H. Some factors modifying the expression of human strength. *Journal of Applied Physiology,* 1961, **16**, 157–163.

ILLOVSKY, J. An experience with group hypnosis in reading disability in primary behavior disorders. *Journal of Genetic Psychology,* 1963, **102**, 61–67.

KLINE, M. V., GUZE, H., & HAGGERTY, A. D. An experimental study of the nature of hypnotic deafness: Effects of delayed speech feedback. *Journal of Clinical and Experimental Hypnosis,* 1954, **2**, 145–156.

KLOPP, K. K. Production of local anesthesia using waking suggestion with the child patient. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 59–62.

LEVITT, E. E., & BRADY, J. P. Muscular endurance under hypnosis and in the motivated waking state. *International Journal of Clinical and Experimental Hypnosis,* 1964, **12**, 21–27.

LONDON, P., & FUHRER, M. Hypnosis, motivation, and performance. *Journal of Personality,* 1961, **29**, 321–333.

LUDWIG, A. M., LYLE, W. H., JR. Tension induction and the hyperalert trance. *Journal of Abnormal and Social Psychology,* 1964, **69**, 70–76.

MALMO, R. B., BOAG, T. J., & RAGINSKY, B. B. Electromyographic study of hypnotic deafness. *Journal of Clinical and Experimental Hypnosis,* 1954, **2**, 305–317.

MARCUSE, F. L. *Hypnosis: Fact and fiction.* Baltimore: Penguin Books, 1959.

MELEI, JANET P., & HILGARD, E. R. Attitudes toward hypnosis, self-predictions, and hypnotic susceptibility. *International Journal of Clinical and Experimental Hypnosis,* 1964, **12**, 99–108.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58**, 277–299.

PARKER, P. D., & BARBER, T. X. "Hypnosis," task motivating instructions, and learning performance. *Journal of Abnormal and Social Psychology,* 1964, **69**, 499–504.

PATTIE, F. A. The genuineness of unilateral deafness produced by hypnosis. *American Journal of Psychology,* 1950, **63**, 84–86.

PATTIE, F. A. Methods of induction, susceptibility of subjects, and criteria of hypnosis. In R. M. Dorcus (Ed.), *Hypnosis and its therapeutic applications.* New York: McGraw-Hill, 1956.

PRONKO, N. H., & HILL, H. A study of differential stimulus function in hypnosis. *Journal of Psychology,* 1949, **27**, 49–53.

ROCK, N. L., & SHIPLEY, T., JR. Ability to "fake" color blindness in the waking state: A control for suggested color blindness under hypnosis. Philadelphia: Department of Psychiatry, Temple University, 1961. (Mimeo)

ROSENHAN, D. L., & TOMKINS, S. S. On preference for hypnosis and hypnotizability. *International Journal of Clinical and Experimental Hypnosis,* 1964, **12**, 109–114.

SALZBERG, H. C. An experimental investigation of the differential effects of hypnotic, post-hypnotic and waking suggestion on learning, with tasks varied in complexity. Unpublished doctoral dissertation, University of Tennessee, 1960. (a)

SALZBERG, H. C. The effects of hypnotic, post-hypnotic and waking suggestion on performance using tasks varied in complexity. *International Journal of Clinical and Experimental Hypnosis,* 1960, **8**, 251–258. (b)

SARBIN, T. R., & MADOW, L. W. Predicting the depth of hypnosis by means of the Rorschach test. *American Journal of Orthopsychiatry,* 1942, **12**, 268–271.

SCHNECK, J. M. (Ed.) *Hypnosis in modern medicine.* (3rd ed.) Springfield, Ill.: Charles C Thomas, 1963.

SEARS, A. B. A comparison of hypnotic and waking learning of the International Morse Code. *Journal* *of Clinical and Experimental Hypnosis,* 1955, **3,** 215–221.

SECTER, I. I. An investigation of hypnotizability as a function of attitude toward hypnosis. *American Journal of Clinical Hypnosis,* 1960, **3,** 75–89.

SECTER, I. I. Personality factors of the MMPI and hypnotizability. *American Journal of Clinical Hypnosis,* 1961, **3,** 185–188.

SHOR, R. E. Physiological effects of painful stimulation during hypnotic analgesia under conditions designed to minimize anxiety. *International Journal of Clinical and Experimental Hypnosis,* 1962, **10,** 183–202.

SHOR, R. E., ORNE, M. T., & O'CONNELL, D. N. Validation and cross-validation of a scale of self-reported personal experiences which predicts hypnotizability. *Journal of Psychology,* 1962, **53,** 55–75.

STUKAT, K. G. *Suggestibility: A factorial and experimental analysis.* Stockholm: Almqvist & Wiksell, 1958.

SUTCLIFFE, J. P. "Credulous" and "skeptical" views of hypnotic phenomena: Experiments in esthesia, hallucination, and delusion. *Journal of Abnormal and Social Psychology,* 1961, **62,** 189–200.

WEITZENHOFFER, A. M. *General techniques of hypnotism.* New York: Grune & Stratton, 1957.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto: Consulting Psychologists Press, 1959.

WEITZENHOFFER, A. M., & SJOBERG, B. M., JR. Suggestibility with and without "induction of hypnosis." *Journal of Nervous and Mental Disease,* 1961, **132,** 204–220.

WEITZENHOFFER, A. M., & WEITZENHOFFER, GENEVA B. Personality and hypnotic susceptibility. *American Journal of Clinical Hypnosis,* 1958, **1,** 79–82.

WELLS, W. R. Expectancy versus performance in hypnosis. *Journal of General Psychology,* 1947, **35,** 99–119.

WHITE, R. W. Prediction of hypnotic susceptibility from a knowledge of subject's attitudes. *Journal of Psychology,* 1937, **3,** 265–277.

WILLEY, R. R. An experimental investigation of the attributes of hypnotizability. Unpublished doctoral dissertation, University of Chicago, 1951.

YATES, A. J. Delayed auditory feedback. *Psychological Bulletin,* 1963, **60,** 213–232.

(Early publication received July 29, 1964)

Volume 70 June 1965 Number 3

Journal of Abnormal Psychology

Editor: HOWARD F. HUNT, *New York State Psychiatric Institute and Columbia University*

Associate Editor: WILLIAM N. THETFORD, *College of Physicians and Surgeons, Columbia University*

CONSULTING EDITORS

THEODORE X. BARBER
Medfield Research Foundation

HAROLD BASOWITZ
New York University

EUGENE I. BURDOCK
New York State Psychiatric Institute and Columbia University

JOHN M. BUTLER
University of Chicago

ARTHUR C. CARR
New York State Psychiatric Institute and Columbia University

LOREN J. CHAPMAN
Southern Illinois University

JOHN H. CUMMING
New York State Department of Mental Hygiene

W. GRANT DAHLSTROM
University of North Carolina

JEROME D. FRANK
Johns Hopkins University

SOL L. GARFIELD
Columbia University

ANN MAGARET GARNER
Lincoln, Nebraska

ISRAEL GOLDIAMOND
Institute for Behavioral Research

HARRISON G. GOUGH
University of California, Berkeley

STARKE R. HATHAWAY
University of Minnesota

RALPH HEINE
University of Chicago

NICHOLAS HOBBS
George Peabody College

LAWRENCE C. KOLB
New York State Psychiatric Institute and Columbia University

SHELDON J. KORCHIN
University of California

WILLIAM G. LAWLOR
Fordham University

ARDIE LUBIN
U. S. Naval Hospital, San Diego

SALVATORE MADDI
University of Chicago

DAVID ROSENTHAL
National Institute of Mental Health

ALAN K. ROSENWALD
Illinois State Psychiatric Institute

SAUL ROSENZWEIG
Washington University, St. Louis

JACQUES RUTSCHMANN
New York State Psychiatric Institute

SEYMOUR SARASON
Yale University

JOSEPH SCHACHTER
New York State Psychiatric Institute and Columbia University

ROY SCHAFER
Yale University

JOHN SHLIEN
University of Chicago

MORRIS I. STEIN
New York University

CHARLES STROTHER
University of Washington

HAROLD L. WILLIAMS
University of Oklahoma School of Medicine

C. L. WINDER
Michigan State University

JOSEPH ZUBIN
New York State Psychiatric Institute and Columbia University

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	FRANCES L. BREWER	ELIZABETH S. REED	ROBERT L. HOOD
Managing Editor	Editorial Assistant	Advertising Manager	Circulation Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 155–164

USE OF REINFORCEMENT AND IMITATION TO REINSTATE VERBAL BEHAVIOR IN MUTE PSYCHOTICS[1]

JAMES A. SHERMAN[2]

University of Washington

Verbal behavior was reinstated in 3 long-term, mute psychotics using reinforcement procedures. For 2 Ss, shaping and fading techniques were used to establish limited verbal repertoires. With a 3rd S, the shaping technique was ineffective; however, a reinforced imitation technique succeeded in developing imitative vocalizations which were then established as verbal responses. For all Ss, the functional role of contingent reinforcement in maintaining the recently developed verbal behavior was demonstrated: when Ss received reinforcement only when they were not responding verbally, their rates of verbal responding showed marked decreases, but recovered when contingent reinforcement for verbal behavior was resumed.

Behavior therapy is a term currently used to label the modification of abnormal human behavior through procedures developed by experimental psychology. In recent years behavior therapy studies have begun to appear with increasing frequency. A critical review of this area has been presented by Sherman (1964).

A major problem in the area of behavior therapy has been the lack of experimentation. Certainly it is worthwhile to provide examples that a certain procedure may be useful for therapeutic purposes. However, it seems more valuable first to demonstrate experimentally that the procedure used was indeed responsible for producing this changed behavior. With individual cases, such an experimental demonstration can be provided by showing that systematic manipulations of the procedure repeatedly result in reliable changes in the problem behavior (Sidman, 1960).

Two examples of successful but nonexperimental behavior-therapy studies are provided by Isaacs, Thomas, and Goldiamond (1960) and by Sherman (1963). In the Isaacs et al. study, a shaping procedure was used to reinstate verbal behavior in two previously mute psychotics. For one subject the reinstated verbal behavior was obtained only in response to questions posed by the experimenter in the experimental situation. The subject apparently did not respond to questions asked by others; nor did he respond on the ward to questions even when asked by the experimenter. In the other case, verbal behavior was reinstated both in the experimental situation and on the ward, but only in response to the experimenter. An attempt was made to generalize this verbal responding to other people by bringing another person into the experimental situation. After a period of a month, the subject began to respond verbally to the second person. He also began to respond to people on the ward when the ward situation was arranged so that his nonverbal requests had no reinforcing consequences.

In the Sherman study, the subject was a mute psychotic who communicated with ward attendants by writing notes. The experimenter refused to read the subject's notes and used a shaping procedure to establish a verbal response. The reinstatement of one verbal response was accompanied by an almost total reinstatement of other verbal behavior in response to the experimenter. In addition, once ward attendants refused to read the subject's notes, he responded verbally to them on the ward.

Unfortunately, in neither study were systematic manipulations of procedure performed to determine their effect on verbalizations. While it appeared highly likely that the reinforcement procedures were responsible for the subjects' verbal behavior, this could not be stated with complete confidence. It could have been, for example, that the extrinsic reinforcement was irrelevant, and that some other unspecified "social relationship" was responsible for the increase in verbalizations. Or, in the Isaacs et al. study, since the subjects were concurrently participating in group therapy, this might have contributed to the verbal reinstatement.

---

[1] This investigation was supported in part by a Public Health Service fellowship (1-F1-MH-20,426-01) from the National Institute of Mental Health. This paper is a portion of the research submitted in partial fulfillment of the requirements for the PhD degree at the University of Washington, 1964. The author is grateful to Donald M. Baer for his advice and encouragement throughout the course of this study.

[2] Now at Walter Reed Army Institute of Research. Reprints may be obtained from the author, Institute for Behavioral Research, 2426 Linden Lane, Silver Spring, Maryland.

The purpose of the present study was (*a*) to use positive reinforcement procedures to reinstate some verbal behavior in long-term, mute psychotics, and (*b*) to manipulate conditions to demonstrate experimentally that the contingent use of reinforcement was essential to this reinstatement. A secondary goal was to evaluate the generalization of the verbal behavior to persons other than the experimenter and to situations other than the experimental one.

METHODS AND RESULTS

The subjects were three hospitalized psychotics at a state hospital.[3] All subjects had been hospitalized continuously for 20 years or longer and all had had long histories of mutism. A complete description of each subject is given below.

General Procedure

In general, three separate procedures were employed to establish verbal behavior in the subjects: shaping, reinforced imitation, and fading. Shaping procedures are used when it is necessary to develop behavior which is not currently exhibited by a subject, either because of the complexity of the behavior or because of the impoverished behavioral repertoire of the subject. In shaping, the experimenter at first reinforces responses which may have little similarity to the desired behavior, but which are clearly within the subject's current behavioral repertoire. The variability of the subject's behavior allows the experimenter to reinforce successively only behaviors which are increasingly similar to the desired behavior. Finally, only the desired behavior is reinforced until it attains suitable strength.

Although imitation has often been cited as having important functions in behavior development, it rarely has been used to modify abnormal human behavior (for one of the few examples see Jones, 1924). In the present study, a reinforced imitation technique was employed with one subject (Subject 3) with whom shaping techniques appeared to be ineffective. An imitative procedure was chosen because a study by Baer and Sherman (1964), with children, demonstrated that the probability of occurrence of an imitative bar-pressing response could be changed by contingent reinforcement, not directly of bar pressing, but of other imitative responses. This study thus suggested that a class of imitative responses could be established such that the strengthening of a few members of the class would result in the strengthening of other members of the class.

The implications of the Baer and Sherman study for the establishment of verbal behavior in a subject seemed clear: if a repertoire of various nonverbal imitative behaviors could be established, it might result in an increased probability of occurrence of imitative vocal and verbal behavior. In the imitative procedure used in the present study, the subject initially was required to imitate various nonverbal behaviors in order to receive reinforcement; gradually the behaviors to be imitated progressed to behaviors associated with vocalizations (e.g., mouth movements), and then to vocalizations and verbalizations.

Fading procedures may be used to change the stimulus conditions in which a behavior is exhibited, without changing the form of the behavior and without repeated extinction trials and consequent errors. The procedure is to introduce the stimulus change very gradually, at a rate which produces no disruption in correct performance. In this manner, an already established behavior may be errorlessly brought under the control of new discriminative stimuli. For example, in programed instruction a verbal response of the subject may be obtained by providing that assortment of textual cues which invariably leads to the emission of the correct response. Then the textual cues are gradually changed so that the subject responds correctly to what was a complex question without the presence of the textual cues which originally controlled the response. Fading procedures may also be used to program, rapidly and with few errors, the learning of stimulus discriminations, where standard discrimination procedures would involve more errors and longer time periods (Moore & Goldiamond, 1964; Terrace, 1963a, 1963b).

In the present study, once subjects were mimicking some of the experimenter's verbalizations, fading procedures were used to establish the subjects' verbal behavior in response to the experimenter's questions.

Once shaping, reinforced imitation, and fading procedures had established stable rates of verbal behavior in the subjects, periods of differential reinforcement for other behavior (DRO) were programed to investigate the role of contingent reinforcement in maintaining the verbalizations. During DRO periods all conditions in the experimental sessions were kept the same as in the preceding sessions, except the reinforcement contingencies. Reinforcement was delivered only when the subjects were not responding verbally, but was delivered in the same amount and at approximately the same rate as previously. Thus, any changes in a subject's rate of verbal behavior during the DRO period could be directly attributed to the changed reinforcement contingencies rather than any other factor in the experimental situation.

---

[3] The author is indebted to Giulio di Furia, Superintendent, Western State Hospital, and John B. Marks, Director, Mental Health Research Institute, Fort Steilacoom, Washington, for their assistance in making space and subjects available for this study.

Following the DRO periods, contingent reinforcement for verbal behavior was resumed until the subjects again showed stable rates of verbalizations. Further reinforcement procedures were then used to expand the subjects' verbal repertoires.

A test of the generalization of the subjects' responding to other people was made by having a second experimenter come into the experimental situation and take over the sessions, both in the presence and absence of the first experimenter. The generalization of the subjects' verbal behavior to other situations was evaluated by their verbal responses to both the first experimenter and the second experimenter on the ward.

The responses of the subjects during sessions were recorded both with a tape recorder and in writing by the experimenter. Checks upon the reliability of the experimenter's recording were not attempted since the responses of the subjects (once they were responding with words) were easily and clearly discriminable. One indication of the reliability of recording was obtained from the generalization test periods when a second experimenter took over the experimental sessions and recorded the responses of the subjects, both in writing and by tape recording. The tapes of these sessions were later scored by the first experimenter. There were no differences between the two records, either in type or in number of words.

Specific Procedures and Results

Since the specific application of the procedures varied among subjects, a detailed description for each subject is required. The experimental room which was used for sessions contained a desk, three chairs, and a tape recorder.

Subject 1

Subject 1 was a 63-year-old man, diagnosed, in 1916, as dementia praecox, hebephrenic type. He had been in the hospital continuously for 47 years, with a history of mutism for 45 of those years. At the time of this study, he was not receiving any medication or participating in psychotherapy. Periodically, when seen on the ward, Subject 1 could be observed walking around mumbling softly to himself. However, all of this mumbling appeared to be nonsensical vocal behavior. In his 45-year history of mutism, Subject 1 had not exhibited any recorded instance of appropriate verbal behavior.

Sessions 1–2. Sessions were held 3 times a week and were approximately ¾ of an hour in length. Initially, candy and cigarettes were used as reinforcers. When a reinforcer was presented, it was accompanied by a statement from the experimenter such as, "Good" or "Very good."

Since the subject did not attend to the experimenter most of the time, and since it was felt that the refinement of later verbal behavior would, to some degree, be dependent upon keeping the subject's attention, the first response reinforced was making eye contact with the experimenter for a 1-second period. By the end of Session 2, the subject was making eye contact with the experimenter an increasing amount of time.

Sessions 3–8. At the start of Session 3, the subject spontaneously emitted a vocalization (a grunt) and was reinforced for it. Thereafter, reinforcement was made contingent upon vocalizations. Vocalizations were defined as any audible sound, including moans, grunts, burps, and coughs. Had words occurred, they too would have been reinforced, of course. However, words never did occur until after direct shaping during the early sessions.

By Session 6 the subject exhibited a low rate of vocalizations (five to eight per session); however, by Session 8 there was no indication of any further increase in rate.

Sessions 9–11. To increase the effectiveness of the reinforcement procedure, portions of the subject's lunch were made contingent upon vocalizations. Each lunch 3 days a week was divided up into approximately 50 portions; and when a vocalization occurred, the experimenter said, "Good," and handed a bite of food to the subject. Concurrently, checks were made of the subject's weight before each session. The subject showed only minor weight changes between sessions; he weighed 150 pounds at the start of the experiment and 151 pounds at the end.

Sessions 12–25. At the beginning of Session 12, the experimenter started to use instructions. A bite of food was held up and the experimenter said, "Say food." At first, any vocalization which temporally followed these instructions was reinforced. Gradually, however, the requirements for reinforcement were changed so that an increasing similarity to the word "food" was required.

Between Sessions 12 and 22, the responses of the subject progressed from indistinguishable mumbles and grunts to drawn out "ōō"[4] sounds, to "ōōd" sounds, and finally, to a distinct "food." When the response "food" occurred, the subject also began to repeat other simple words said by the experimenter, such as "water," "pie," and "Jello." By Session 25, in addition to the word "food," the subject had repeated 12 different words. However, the experimenter's control over the subject's responding was weak. Frequently, the subject did not respond when the experimenter held up a bite of food and said, "Say food."

Sessions 26–43. To increase the experimenter's control over the subject's verbal responding, all three of the subject's meals for 3 consecutive days were made contingent upon verbal behavior. The experimenter continued to use the instructions, "Say food," while holding up a bite of food. If the subject responded, he was reinforced. If he did not respond within 10 seconds, the experimenter silently read a book for 1 minute before starting the next trial.

By Session 43 the experimenter had good control over the subject's verbal responding. Almost every

---

[4] Standard dictionary phonetic symbols are used to indicate pronunciation.

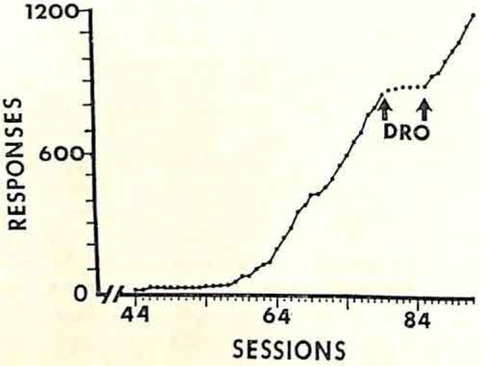


FIG. 1. A cumulative record of the number of verbal responses per session for *S*1 under reinforcement for responses and DRO conditions.

time the experimenter used the instructions, the subject responded, "Food." In addition, the subject consistently repeated any one of approximately 20 words (such as "meat" or "pie") the experimenter might use instead of the word "food." However, the subject did not respond with a nonimitative word. For example, when the experimenter held up a bite of food and said, "Say food," the subject responded, "Food." If on the next trial, the experimenter held up a bite of food and asked, "What is this?" the subject did not respond.

Sessions 44–79. To obtain the word "food" from the subject when the experimenter asked, "What is this?" a fading procedure was used. With the fading procedure, the experimenter continued to hold up a bite of food each time and to deliver instructions to the subject. The behavior of the subject—that is, saying "Food"—was maintained with reinforcement while the instructions to the subject were gradually changed in the following steps: (*a*) "Say food"; (*b*) "Say foo_"; (*c*) "Say f______"; (*d*) "What is this? Say f______"; (*e*) "What is this? Say ______"; (*f*) "What is this?"

By Session 79 the subject was saying the word "food" with great regularity as soon as the experimenter held up the bite of food. In fact, between Sessions 72 and 79 the subject was responding regularly enough so that he consistently received almost all of every experimentally controlled meal. In addition to the word "food," the subject occasionally named specific foods. When he did so, he was reinforced. Nevertheless, approximately 95% of the words he emitted were the word "food."

Sessions 80–85. To determine whether contingent reinforcement was essential in maintaining the subject's verbal behavior, a period of DRO was programed. Now, if the subject responded verbally when the bite of food was held up, he did not receive the food reinforcement until 30 seconds after his response (or if he responded more than once, 30 seconds after his last response). If the subject did not respond, he received the food 30 seconds after it was held up. Thus, the subject received the food (accompanied by "Good" from the experimenter) whether he responded or not. All other conditions were kept the same as in the immediately preceding sessions. The experimenter continued to hold up the bite of food and ask, "What is this?" at the same rate, and the food was delivered at approximately the same rate and in the same amount as previously.

Figure 1 shows the effect of DRO sessions on the subject's previously stable rate of verbal responding. This graph is a cumulative plot of the total number of nonimitative words emitted by the subject each session (mainly "food"). As can be seen from the figure, the DRO period (dotted line) resulted in a marked and stable decrease in verbal responding.

Sessions 86–92. Contingent food reinforcement for verbalizations was resumed. However, now when the experimenter held up the bite of food and asked, "What is this?" there was no response from the subject to be reinforced. Therefore, at the start of Session 86, the experimenter returned to an earlier step in the reinforcement procedure. The experimenter held up the bite of food and said, "Say food." The subject responded and was reinforced. After this had been repeated three times, a brief fading procedure, similar to that used during earlier sessions, was used until, after about 15 minutes, the subject said "Food" when the experimenter merely held up the bite of food. Thereafter, verbal behavior was simply reinforced as before.

As can be seen from Figure 1, the reinstitution of reinforcement for verbal behavior, plus the brief fading procedure, resulted in a recovery of the subject's former rate of nonimitative verbalizations.

Sessions 93–122. Up to this time the subject's verbal behavior consisted primarily of one word, the word "food." Next, an attempt was made to see how long it would take to establish 10 new words in his repertoire. For this, 10 picture cards were employed. Each of these cards contained a picture of some common object or animal such as a ship, a dog, a rabbit, or some grapes. To establish these 10 new verbal responses, food reinforcement was continued and a fading procedure was used. First, the experimenter pointed to the picture of the dog and said, "Say dog." When the subject repeated the word, he was reinforced. Gradually the word "dog" was faded out from the experimenter's instructions until the subject was responding with "Dog" when the experimenter pointed to the card and said, "What is this?" A similar fading technique was used for the other 9 cards, and by Session 122 the subject was responding appropriately and reliably to each of the 10 cards.

Total verbal behavior. At the end of the experimental sessions, the subject's verbal repertoire consisted of approximately 30 words which could be reliably obtained. The subject would say "Food," "Candy," "Cigarette," name the 10 picture cards, read his name and read the numbers from 1 to 20 printed on a card. The name and number reading responses appeared to be indirect effects of the reinforcement of other verbal behavior, since these responses were never directly shaped.

Six months after the experimental sessions had been terminated, the subject was tested to see whether his verbal behavior could still be obtained. All the verbal behavior which the subject exhibited at the end of the experimental sessions was still obtainable after this 6-month delay.

Subject 2

Subject 2 was a 45-year-old man diagnosed as schizophrenic. He had been hospitalized continuously for 20 years and had had a history of mutism for 16 of those years. During this study Subject 2 was not taking part in psychotherapy, but was receiving medication (tofranil, 25 milligrams, three times a day).

From ward reports and questioning of the attendants, it appeared that Subject 2 had emitted approximately six words in the 6-month period prior to the study. All of these words were either "Yes" or "No" in response to the attendants' questions.

Sessions 1–4. Candy was used as the reinforcer, accompanied by "Good" or "Very good" from the experimenter. Sessions were held two times a week and were approximately ¾ of an hour in length.

The word "candy" was shaped through reinforcement of the subject's increasing approximations to the word when the experimenter held up a piece of candy and asked, "What is this?" By Session 4 "Candy" could be reliably obtained from the subject whenever the experimenter held a piece of candy up. In addition, during these early sessions the experimenter periodically asked the subject various conversational-type questions such as, "What is your name?" "What do you like to read?" "What did you do yesterday?" and the like. At first, any type of vocalization following one of these questions was reinforced; gradually, however, the subject was required to say at least one intelligible word in order to receive reinforcement. During Session 4 the subject responded for the first time with distinct words to some of the conversational questions.

Sessions 5–17. Since the subject was responding perfectly when candy was held up, beginning with Session 5, the experimenter began to ask the subject to name various objects in the experimental room such as a chair, a table, a pencil, and so forth. Each time the experimenter would point to the object and say, "What is that?" In addition, mixed in with object questions, the experimenter continued to ask the conversation questions. For both types of questions, if the subject responded appropriately he was reinforced; if not, the experimenter silently read a book for 1 minute before asking the next question.

Figure 2 shows the results of this procedure. This figure is a plot of the percentage of each type of question, object and conversation, to which the subject replied each session. A reply to a question consisted of one or more intelligible words that were appropriate to the question being asked. As can be seen from Figure 2, the subject exhibited an increased percentage of questions answered for both object and conversation questions.

Sessions 18–35. To determine whether contingent reinforcement was functional in increasing the subject's verbal behavior, several DRO periods were programed. In these periods the experimenter continued to ask both object and conversation questions at the same rate as previously. During DRO conditions, if the subject responded, the experimenter waited until 30 seconds after the response and then reinforced the subject with candy and a verbal "Good." If the subject did not respond, he received reinforcement 30 seconds after the experimenter's question. All of the object questions and the great majority of the conversation questions used were ones that had been asked in the immediately preceding sessions. (The differences in the experimenter's conversation questions between sessions were due to differences in the subject's answers which necessitated slightly different follow-up questions.)

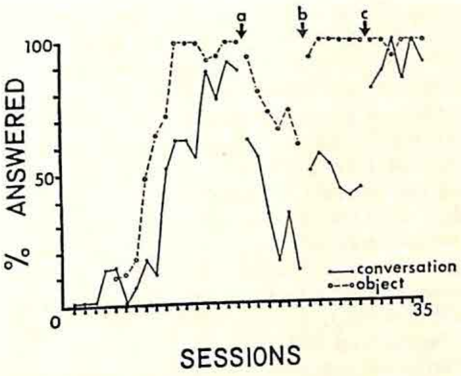


FIG. 2. A record of the percent of verbal responses per session for *S2* to two types of questions under reinforcement for responses and DRO conditions. (At (a) answers to both types of questions were put under the DRO condition; at (b) answers to object questions were reinforced, with answers to conversation questions kept under the DRO condition; at (c) answers to both types of questions were reinforced.)

In Sessions 18–23, responses to both types of questions were under the DRO conditions. Figure 2 shows the effect upon the percentage of questions answered: answers to both object and conversation questions showed a marked decrease, with the conversation answers showing the greater decrease.

During Sessions 24–29, contingent reinforcement for answers to object questions was resumed, while DRO procedures were left in effect following answers to conversation questions. With this procedure, percentage of answers to object questions rose to its former rate, while percentage of answers to conversation questions showed an initial rise, although not to its former rate, and began to drop.

In Sessions 30–35, contingent reinforcement was resumed for answers to conversation questions, with contingent reinforcement for answers to object questions kept in effect. The results, plotted in Figure 2, show that responding to conversation questions rose markedly, while responding to object questions remained high.

Sessions 36–44. Up to this time the subject had responded to conversation questions primarily with one or two word answers. To increase the number of words spoken per answer, object questions were discontinued and a fading procedure was used for conversation questions. For example, when the experimenter would ask the subject, "What did you do yesterday?" the subject might respond "Walk." The experimenter would then reply, "No, make a sentence. Say, *I went for a walk.*" If the subject repeated the sentence, he was reinforced. Each time the subject responded to the experimenter's questions with one or two words, the experimenter made a sentence out of the subject's response and told the subject to repeat the sentence. Gradually, the verbal cues supplied by the experimenter were faded out until, when the subject responded with one or two words to a question, the experimenter only had to say, "Make a sentence," to obtain a sentence or complete phrase from the subject. Finally, by Session 44 even these last instructions were dropped. The subject was regularly responding to common conversational questions with full-sentence answers, averaging about five to six words per answer.

Subject 3

Subject 3 was a 61-year-old woman, diagnosed (in 1926) as dementia praecox, simple type. She had been hospitalized continuously for 37 years and had a history of complete mutism for 33 of those years. During this study Subject 3 was not participating in psychotherapy, but was receiving medication (equanil, 400 milligrams, two times a day).

From the ward reports and questioning of the attendants, it appeared that the subject had not exhibited any verbal behavior in the last 33 years. In fact, the attendants reported that they had never heard the subject emit any type of vocalization.

Sessions 1–10. Initially, candy acompanied by "Good" or "Very good" from the experimenter was used as the reinforcement. Sessions were held three times a week and were approximately ¾ of an hour in length.

Since it appeared that the subject did not exhibit a rate of vocalizations, the first response chosen for reinforcement was lip movement. It was hoped that as lip movements increased in probability of occurrence, vocalizations also would occur. At first, reinforcement was presented whenever very slight lip movements occurred; later, reinforcement was contingent upon a distinct parting of the lips. Once a steady rate of parting the lips was exhibited, this response was put on a variable ratio schedule of reinforcement (VR 3) to further increase its rate.

Sessions 10–23. During the last part of Session 10, the subject emitted a vocalization for the first time (a low moan), and was reinforced for it. Thereafter, the experimenter began to use instructions of "Say candy" while holding up a piece of candy. Reinforcement was presented contingent upon any type of sound which followed these instructions, including soft vocalizations, burps, coughs, and audible sighs. If the subject did not respond to the instructions within 20 seconds, the experimenter silently read a book for 1 minute before starting the next trial.

Between Sessions 10 and 23 the subject exhibited a number of sounds; however, the rate of these sounds was extremely low. They also were of very low intensity, so low that the experimenter had difficulty in hearing them even though he sat with his ear approximately 8 inches from the subject's mouth.

Sessions 24–35. In an attempt to increase the effectiveness of the reinforcement procedure, the subject's meals were used as reinforcers. For 3 consecutive days each week the subject's breakfast, lunch, and dinner were each divided up into approximately 60 portions. These portions were presented to the subject, accompanied by "Good" from the experimenter, whenever she emitted a sound following the experimenter's instructions, "Say food."

As in the case of Subject 1, with the introduction of food reinforcement, weight checks were taken before each session. It was found that as experimental sessions progressed, the subject showed a slight decrease in her weight. However, since she was quite overweight at the start of the experimental procedure, this was not of great concern. (The subject was 5 feet, 1 inch tall, and at the start of the experiment weighed 178 pounds; at the end of the experiment she weighed 168 pounds.)

The introduction of food reinforcement failed to increase substantially the subject's low rate of emitting sounds, and by Session 35 it appeared doubtful that continued application of the same procedure would be useful, at least without better methods of detecting faint sounds emitted by the subject.

Sessions 36–57. In order to obtain a rate of sounds from the subject, the imitative procedure described earlier was devised. Again, the purpose of using an imitative procedure was to attempt to establish a functional class of imitative behavior, such that strengthening imitative nonverbal responses would result in the strengthening of imitative vocal and verbal responses.

As a starting point for the imitative procedure, various nonverbal behaviors of the subject which were under the experimenter's verbal control were used. These responses were those of standing up, sitting down, picking up a spoon, and touching the table. Initially, the experimenter said to the subject, "Do this. Stand up." The experimenter then stood up. If the subject performed the response, she was reinforced with a bite of food and "Good" from the experimenter. If she did not perform the response, the experimenter sat down and read for 1 minute before starting the next trial. Gradually, the experimenter changed two aspects of the procedure: first, the verbal description of the action was dropped until the experimenter merely said, "Do this," and performed the response to be imitated; second, the responses exhibited by the experimenter to be imitated were gradually changed until they became more and more similar to vocal behavior. At first, the responses exhibited by the experimenter to be imitated were those of standing up, sitting down, picking up a spoon, and touching the table. Next, other responses such as nodding the head, looking up at the ceiling, and opening the mouth were introduced. After this, responses of blowing out a match and clearing the throat were included.

By Session 40 the subject reliably imitated any of the above responses when the experimenter said, "Do this," and performed the response. At this point the experimenter introduced a vocalization to be imitated —an "äää" sound. Not only did the subject fail to imitate this sound, but following its introduction, she did not imitate any of the responses which had been introduced during the previous two sessions.

Sessions 41 and 42 were spent re-establishing the imitative repertoire the subject had exhibited prior to Session 40. At Session 43, instead of again moving to vocalizations, several intermediate responses were introduced. These responses progressed from blowing (without the match present), to making a hissing sound, to an aspirate "h" sound, to a "ppph" sound, and finally, to an "ààà" sound. By Session 45 these responses were under good imitative control and further sounds were introduced without difficulty.

By Session 48 the subject was imitating the sounds of "fff," "ōō," and "da," which, when combined, make up the word "food." Since the subject did not, at this point, imitate the entire word, Sessions 49 to 57 were spent chaining these sounds together to make a functional verbal unit. Initially, the experimenter said to the subject, "Say fff (pause while the subject responded) Say ōō." When the subject imitated first "fff" and then "ōō," she was reinforced. Gradually, the instructions were changed to, "Say fff (pause while the subject responded) ōō," and then, "Say fffōō." In a similar manner, "da" was added onto the previously built chain of "fffōō" until, at Session 57, the subject consistently imitated the entire word "food." In addition, it was noted at this point that the subject would consistently imitate many other simple words said by the experimenter.

Sessions 58–78. Since the subject at this point was simply repeating words said by the experimenter, the next step was to have her emit "Food" when the experimenter held up a bite of food and asked, "What is this?" To establish this response, a fading procedure identical with that used for Subject 1 was used, and by Session 66 the subject was responding regularly with "Food" when the experimenter asked, "What is this?" As the reinforcement procedure continued, other words began to appear in the subject's repertoire. She no longer always answered "Food" in response to the experimenter's question, but on occasion named specific foods. These words which appeared were ones which had never been directly shaped by the experimenter. However, if they were appropriate to the kind of food being held up, the subject was reinforced.

From Sessions 65 to 78 the subject responded regularly enough to receive almost all of each experimentally controlled meal. Figure 3 is a plot of this responding. It is a cumulative plot of the total number of responses emitted by the subject each session in response to the experimenter's question, "What is this?" Each of these responses consisted either of "Food" or of the name of a food. In general, as sessions progressed the response "Food" became less frequent and, with increasing regularity, the subject responded to the experimenter's question with the name of a specific type of food. Thus, the plot of the responses for Sessions 68 to 78 in Figure 3 represents the combined totals of approximately 40 different words.

Sessions 78–85. Since a stable rate of verbal responding was established, DRO procedures were programed to examine the role of reinforcement in maintaining the verbal behavior. This DRO period was put into effect during the second half of Session 78. While it was in effect the experimenter continued to hold up bites of food and ask, "What is this?" at the same rate as he had before. If the subject responded, she received reinforcement 30 seconds after her response (or if she responded more than once, 30 seconds after her last response). If she did not respond, she received reinforcement 30 seconds after the experimenter said, "What is this?"

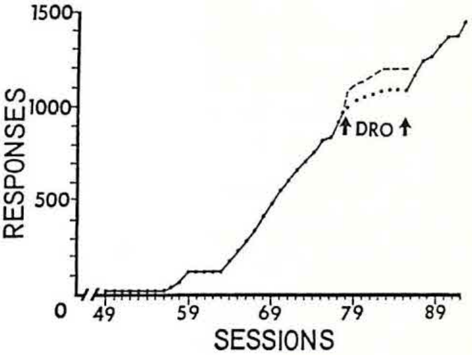


FIG. 3. A cumulative record of the number of verbal responses per session for *S3* under reinforcement for responses and DRO conditions.

Figure 3 shows the results of the DRO period (dotted lines) upon the subject's rate of verbal responding. The two dotted lines shown in the figure representing the DRO period were necessary because of the subject's type of response under this condition. The upper dotted line shows the total number of responses, including repeated responses, emitted by the subject each session. At first, when the experimenter held up the bite of food and asked, "What is this?" the subject might respond "Meat." When reinforcement was not forthcoming, the subject again responded "Meat," and then again and again. On some trials the subject responded over 20 consecutive times before pausing long enough (30 seconds) to receive reinforcement. Thus, because of the repeated responses, the upper dotted line shows an initial increase over the previous rate of verbal responding. These repeated responses, however, soon ceased to be emitted. The lower dotted line of Figure 3 represents the total number of trials responded to with one or more appropriate words per session during the DRO period. In spite of the differences in plotting, both lines show essentially the same characteristic: during the DRO period the subject's rate of verbal responding gradually decreased.

Sessions 86–92. Contingent reinforcement for verbalizations was resumed. Initially, the experimenter had to name several foods in order to obtain a response from the subject to reinforce. After several trials of this, however, the subject again began to emit naming responses when the experimenter held up a bite of food and asked, "What is this?"

The marked increase in the rate of verbal behavior (shown in Figure 3) following the resumption of contingent reinforcement, further demonstrated the functional role that reinforcement played in reproducing the subject's verbal behavior.

Sessions 93–99. These sessions were spent expanding the subject's verbal repertoire. A number of objects and pictures of objects were introduced into the experimental sessions, and the subject was required to name them in order to receive reinforcement. For the most part, the subject named these objects immediately, and it was not required for the experimenter to name them first. In addition, a limited variety of conversational questions and comments were included in each session. These questions and statements were ones such as, "What is your name?" "What ward are you from?" "How are you today?" "Hello," and "Goodby." If the subject responded to these appropriately, she was reinforced.

By the end of Session 99 the subject was consistently naming scores of objects. She also was responding regularly with appropriate replies to the simple questions and statements made by the experimenter, although her replies consisted only of one or two words.

Generalization Tests

Responding to other people. At the end of the experimental procedure, a test of the generalization of the subjects' responding to other people was made by having a second experimenter (a male for Subject 1 and a female for Subject 2 and Subject 3) come into the experimental situation and take over the sessions. The second experimenters were given instructions to duplicate as much as possible the behavior that the first experimenter exhibited during the last experimental session with each subject (including reinforcement for subjects' verbalizations).

The responses of all three subjects in response to the second experimenter were quite similar to their verbal behavior toward the first experimenter. Subject 1 continued to name the 10 picture cards accurately; Subject 2 responded with full sentence or phrase answers to conversation questions and named all the objects he was asked to name; Subject 3 named objects and gave one-word answers to certain conversational questions and statements. These responses from the subjects were obtained both with the first experimenter present and with him absent from the room. Thus, the newly reinstated verbal behavior of the subjects was not limited to a specific person.

Responding in other situations. During the course of the study it had been noted that the subjects' verbal behavior was not limited to the experimental room. In fact, each of the subjects exhibited all the verbal behavior on the ward that they exhibited in the experimental room as long as at least a partial reinforcement schedule was maintained. For example, when Subject 1 was repeating simple words in the experimental room he also repeated these words on the ward; when he was naming picture cards in the experimental room, he named them on the ward. The same type of effect was noted for Subject 2 and Subject 3. With all three subjects, this verbal behavior on the ward was exhibited both toward the first experimenter and toward the second experimenter.

In spite of the verbal behavior which was exhibited toward both the experimenters on the ward, the subjects' responding did not seem to generalize greatly to ward attendants. Subject 1 apparently did not ever exhibit any verbal behavior toward attendants on the ward after the experiment was completed. With Subject 2, ward attendants reported that he responded verbally to their questions several times a week, whereas he had exhibited only about six words on the ward in the 6-month period prior to the study. Subject 3, on at least one occasion, exhibited verbal behavior in the presence of an attendant.

Part of the reason for the failure of subjects' verbal behavior to generalize to ward attendants may have been due to a lack of reinforcement for verbalizations. When the subjects responded verbally to the experimenter on the ward, at least a partial reinforcement schedule had to be in effect to maintain this responding. If reinforcement was completely discontinued, the subjects quickly stopped responding. Nevertheless, it seems possible to overcome this rapid cessation of responding by scheduling the reinforcements. At one point with Subject 3, the experimenter started to schedule the reinforcements, gradually increasing to a variable ratio four, with no noticeable decrease in the subject's responsiveness. Thus, it would seem possible that further scheduling might result in sustained verbal behavior with only an occasional reinforcer delivered. Also, perhaps the use of ward attendants to deliver reinforcers for subjects' verbalizations would have increased their responding to the attendants.

DISCUSSION

With the three subjects in this study, an increase in verbal behavior was obtained through the use of reinforcement procedures. In the case of Subject 1, shaping and fading techniques were employed to establish a verbal response, which was then maintained with reinforcement. For Subject 2 a shaping technique was used initially to establish a verbal response. Thereafter, reinforcement was used to maintain and expand the verbal repertoire exhibited by the subject in response to the experimenter's questions. With Subject 3 a shaping technique appeared to be ineffective in establishing behaviors approximating verbal behavior. However, a reinforced imitative technique succeeded in developing imitative vocalizations, which were then chained together to form an imitative verbal unit. Through the use of fading, this imitative verbal unit was established as a verbal response.

For all subjects, once a rate of verbal responding had been established, DRO procedures were put into effect. The results of the DRO periods demonstrated the functional role that contingent reinforcement played in maintaining the subjects' rate of verbal responding, since the rates of all three subjects showed marked decreases under the DRO condition. By extension, these results imply the functional role that reinforcement played in producing the subjects' verbal behavior. In addition, when contingent reinforcement for verbalizations was resumed following the DRO periods, the rates of all three subjects showed marked increases. This provides further evidence regarding the functional role of contingent reinforcement in re-establishing the subjects' verbal behavior.

The intrasubject reversals of the rate of verbal behavior described for Subject 1, Subject 2, and Subject 3 show the reliability of the results for each case. That is, for each of these three cases they demonstrate that the increases in verbal behavior which were obtained were not due to unknown or chance factors, but were due specifically to the contingent use of reinforcement.

Information about the generality of the present results, that is, the typical applicability of positive reinforcement procedures to the reinstatement of behavioral repertoires in disturbed subjects, is provided by an accumulation of many individual cases in which the functional efficacy of such procedures has been demonstrated for each case (Sidman, 1960). Thus, information on generality is provided by the results of the intersubject replications of the use of positive reinforcement procedures with Subject 1, Subject 2, and Subject 3. With each of these subjects, positive reinforcement procedures resulted in an increase in verbal behavior.

Further information on generality is suggested by previous nonexperimental studies which have used positive reinforcement to reinstate behavioral repertoires in hospitalized patients (Ayllon, 1963; Bachrach, Erwin, & Mohr, in press; Brady & Lind, 1961; Isaacs et al., 1960; Sherman, 1963). In addition, there have been several studies which have used positive reinforcement procedures to develop new behavioral repertoires which previously were lacking in subjects (Fuller, 1949; Wolf, Risley, & Mees, 1964).

The similarity of the results reported in the studies cited above lends generality to the findings of the present study; in each of these studies the use of positive reinforcement procedures has been followed by the development of some appropriate behavior. Thus, it can be stated with increased confidence that positive reinforcement procedures can have wide applicability to problems of instatement and reinstatement of normal behavioral repertoires in humans.

One further similarity between the results of the present study and those obtained by Isaacs et al. and by Sherman was the effect of reinstatement of one verbal response upon other verbal responses. For both subjects in the Isaacs et al. study, the reinstatement of the first verbal response was accompanied by the reappearance of many other verbal responses. In the Sherman study the reinstatement of one word was accompanied by an almost total reinstatement of all other verbal behavior. The same type of effect was noted for all three subjects in the present study.

In the cases of both Subject 1 and Subject 3, the establishment of the mimicking response "Food" was acompanied by mimicking responses of other simple words said by the experimenter, even though these latter verbal responses had never been directly shaped. During later sessions with both of these subjects, the establishment of "Food" in response to the experimenter's question, "What is this?" was accompanied by unshaped naming responses of specific types of food.

With Subject 2 there are two observations which are relevant to this point. The first is that during the earlier sessions when he exhibited a naming response to candy, he also began to answer some conversational questions. The second was noted during the subject's DRO periods (see Figure 2). When the DRO condition was first instituted, percentage of responses to both object and conversation questions dropped. Then, when contingent reinforcement was resumed for responses to object questions, the percentage of responses to these questions rose. However, the percentage of responses to conversation questions also rose even though they were still under the DRO condition and were not being reinforced.

Thus, the results of the present study, of Isaacs et al., and of Sherman strongly indicate that, for such subjects, verbal behavior is a functional class: variables which produce changes in one member of the class thereby affect the other members of the class similarly.

There are two main restrictions relating to the generality of the results obtained in the present study. The first is the wide variation among subjects in the amount of verbal behavior obtained following the reinforcement procedures. In the case of both Subject 1 and Subject 3 a quite limited verbal repertoire was reinstated, while with Subject 2 a fairly wide verbal repertoire was re-established. It is difficult to account for the wide differences in amount of reinstatement among subjects through an analysis of the experimental procedures which were used for them.

The second restriction on the generality of the results of this study is the apparent ineffectiveness of shaping procedures to establish vocal behavior in the case of Subject 3. However, the primary reason for this ineffectiveness may have been a result of the difficulty of response detection rather than a lack of applicability of the technique. In shaping, the experimenter must be sensitive to slight differences in the topography of the behavior being shaped. During the early sessions with Subject 3, her vocalizations were of such low intensity that it was difficult to detect their presence; thus reinforcement was likely to be unreliable. One technique which might have alleviated this problem would have been the use of a throat microphone and an amplifier by which these low intensity sounds could be made clearly audible.

Nevertheless, the reinforced imitative technique which was developed for Subject 3 seemed to provide an adequate method of establishing a reliable vocal mimicking repertoire as a starting point from which verbal behavior could be developed. This, after all, is still an example of programing operant behavior with positive reinforcement; it is just a different program.

It seems possible that the type of imitative technique used for Subject 3 could have rather extensive applications. For example, a common behavioral deficit of both autistic and retarded children is their lack of adequate verbal behavior. In cases where these children exhibit vocal mimicking behavior, the use of reinforcement procedures may result in rather rapid progress in verbal training. However, if the children do not exhibit any type of vocal mimicking, or have an extremely low rate of vocalizations, the shaping of verbal behavior may be an arduous process even with relatively powerful reinforcers (Wolf et al., 1964). In these latter cases an imitative procedure similar to that used for Subject 3 may well be useful, in that the development of nonverbal imitative behaviors and subsequent progression toward imitative vocalizations and verbal behavior may be relatively more efficient than the direct shaping of vocalizations and subsequent discrimination training necessary to bring them under the control of other people's verbal behavior.

REFERENCES

AYLLON, T. Intensive treatment of psychotic behavior by stimulus satiation and food reinforcement. *Behaviour Research and Therapy*, 1963, **1**, 53–61.

BACHRACH, A. J., ERWIN, W. J., & MOHR, J. P. The control of eating behavior in an anorexic by operant conditioning techniques. In L. Ullmann & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart, and Winston, in press.

BAER, D. M., & SHERMAN, J. A. Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology*, 1964, **1**, 37–49.

BRADY, J. P., & LIND, D. L. Experimental analysis of hysterical blindness. *Archives of General Psychiatry*, 1961, **4**, 331–339.

FULLER, P. R. Operant conditioning of a vegetative human organism. *American Journal of Psychology*, 1949, **62**, 587–590.

ISAACS, W., THOMAS, J., & GOLDIAMOND, I. Application of operant conditioning to reinstate verbal behavior in psychotics. *Journal of Speech and Hearing Disorders*, 1960, **25**, 8–12.

JONES, MARY C. A laboratory study of fear: The case of Peter. *Pedagogical Seminary and Journal of Genetic Psychology*, 1924, **31**, 308–315.

MOORE, R., & GOLDIAMOND, I. Errorless establishment of visual discrimination using fading procedures. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 269–272.

SHERMAN, J. A. Reinstatement of verbal behavior in a psychotic by reinforcement methods. *Journal of Speech and Hearing Disorders*, 1963, **28**, 398–401.

SHERMAN, J. A. Use of reinforcement and imitation to reinstate verbal behavior in mute psychotics. Unpublished doctoral dissertation, University of Washington, 1964.

SIDMAN, M. *Tactics of scientific research*. New York: Basic Books, 1960.

TERRACE, H. S. Discrimination learning with and without "errors." *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 1–27. (a)

TERRACE, H. S. Errorless transfer of a discrimination across two continua. *Journal of the Experimental Analysis of Behavior*, 1963, **6**, 223–232. (b)

WOLF, M. M., RISLEY, T., & MEES, H. Application of operant conditioning procedures to the behaviour problems of an autistic child. *Behaviour Research and Therapy*, 1964, **1**, 305–312.

(Received July 24, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 165–168

PUPIL RESPONSE OF HETERO- AND HOMOSEXUAL MALES TO PICTURES OF MEN AND WOMEN:

A PILOT STUDY[1]

ECKHARD H. HESS, ALLAN L. SELTZER, AND JOHN M. SHLIEN
University of Chicago

The pupil response of each individual in a group of heterosexual males was greater when looking at pictures of women than when looking at pictures of men. Homosexual male Ss responded in the opposite direction. Measurement of changes in pupil size permitted clear-cut discrimination between the 2 groups.

Change in the size of the pupil of the human eye has been reported to vary with a subject's interest in various pictorial stimuli (Hess & Polt, 1960). Male subjects had a larger pupil while looking at pictures of women than when looking at pictures of men. The reverse was true for female subjects: they had larger pupils looking at men. Unpublished work with a large number of subjects has continued to substantiate the finding of this difference between the sexes.

If this difference in pupil response is truly a reflection of interest in the male or female figure as a sexual object then homosexuals would be expected to show a larger pupil response to pictures of their own sex. In the course of our work a few subjects have given a larger response to pictures of their own sex; as measured by pupil size, same-sex pictures seemed more interesting to them. Review of these anomalous cases increased the plausibility of the idea that this same-sex response might be typical of homosexuals. The present report, a pilot study of a small group of overt male homosexuals, strongly supports that hypothesis.

METHOD

Subjects

Ten young adult male subjects were tested. Five of these, students or workers in our laboratory—the heterosexual group—were well known to us over a period of several years. Their sexual outlet was judged to be exclusively heterosexual. The other five were known, through observation, interview, and in every case by their own voluntary admission to one of the authors who had gained their trust, to have overt homosexuality as their sole or primary sexual outlet. All 10 were of roughly the same age (between 24 and 34 years), same education (all but one were graduate students), and same social level. None was hospitalized or in therapy.

Procedure and Apparatus

In a dimly-lit room, a subject was seated before a viewing aperture, fitted with a headrest, which was inserted in a large plywood panel. The panel concealed the working of the apparatus from the subject. Resting his head against the aperture, the subject faced a rear-projection screen, set in an otherwise black box, at a distance of 2½ feet from his eyes. A 35-millimeter slide projector behind this screen projected a 9 × 12-inch picture onto it. Changing of slides was controlled by the experimenter from his position behind the panel where he also operated a concealed 16-millimeter camera fitted with a frame counter. As the slides were being viewed a half-silvered mirror placed at a 45-degree angle across the subject's line of vision permitted unobtrusive filming of the eye, at the rate of two frames per second. Illumination for this photography was furnished by a 100-watt bulb on rheostat control.

Stimuli

Fifteen picture slides, representations of the human figure, were shown in the following order:

Slide Content	Scoring Category
A. Painting, cubist, five figures	Art
B. Painting, realistic, crucifixion	Art
C. Painting, two nude males	Male
D. Painting, reclining female nude	Female
E. Photograph, nude man, head and upper torso	Male
F. Painting, seated nude female, rear view	Female
G. Painting, sailor, nude upper torso	Male
H. Painting, nude male and nude female	Art
I. Photograph, nude female torso	Female
J. Photograph, nude man, rear view	Male

---

[1] This research was supported in part by a grant from Social Sciences Research Committee of the University of Chicago and in part by Interpublic, New York, New York.

Slide Content	Scoring Category
K. Painting, nude female, head and upper torso	Female
L. Painting, two partly clothed males	Male
M. Painting, nude female, head and torso	Female
N. Painting, abstract, three figures	Art
O. Painting, cubist, three figures	Art

The presentation of each of these stimulus pictures was preceded by the presentation of a medium gray "control" slide. The total sequence was 30 slides in this order: Control A, Stimulus A, Control B, Stimulus B, etc., each shown for 10 seconds, with a total viewing time of 5 minutes for the entire sequence.

From the list of slides it can be seen that five were scored as being pictures of females and five were scored as pictures of males. The "male" pictures (C, E, G, J, and L), considered to be the homosexual equivalent of pinups, were culled from physique magazines and were generally more crude artistically than the pictures of females. These latter (D, F, I, K, and M) represented a rather lush concept of the female figure: for example, "D" was a Titian "Venus," "K" an Ingres "Odalisque."

The five "art" slides (A, B, H, N, and O) ranged in style and period from a Michelangelo to a Picasso. None of these was a clearly male or clearly "female" picture; the abstracts (A, N, and O) were ambiguous sexually, "H" showed both sexes, "B" had a strong religious connotation. This group of slides was included in the series for several reasons. Firstly, it was deemed desirable to place the sexual pictures in an artistic setting to reduce the threat to some subjects that might inhere in the obviously sexual material. Secondly, an abnormally high response is frequently given to the first stimulus shown to a subject. By placing art slides "A" and "B" first in the sequence, the male and the female slides, which were of major interest, were protected from this artifact. Thirdly, homosexuals are often thought to have artistic interests and, indeed, most of the homosexuals in this study did verbally indicate such interests. It was useful, therefore, to include a group of slides which would permit appraisal of response to the artistic quality of pictures separate from their representation of sexual objects. Such a separation of pictorial content from its artistic mode of expression appears feasible since (*a*) the homosexuals, as a group, showed a high response to the artistically good but sexually ambiguous art slides but (*b*) they also showed a high response to the artistically crude male pictures yet (*c*) they showed a low response to the artistically good female pictures. Thus, in addition to the use made of it in this report, the data point also to the potential value of the pupil technique in esthetics research.

Measurement and Scoring

The processed 16-millimeter film was projected, frame by frame, onto the underside of an opal-glass insert in a table, to a magnification of approximately 20 times. The diameter of the pupil in each frame was measured with a millimeter rule and recorded, giving a set of 20 measurements for each control presentation and a set of 20 for each stimulus. Averages were then computed for each stimulus set and for each preceding control set. In order to compare average pupil size during viewing of a picture to the pupil size during the preceding control this method was used: for each control-stimulus pair the percentage of increase or decrease in average pupil size was computed by dividing the difference between stimulus average and control average by the control average. A positive percentage indicated a larger pupil size when the subject was viewing the stimulus than when he viewed the preceding control. A negative percentage meant a smaller average pupil size during stimulus viewing. For each subject, the five percentages of his response to each of the male pictures (C, E, G, J, and L) were added together to give his "response to 'male' picture" score (Table 1, first column). The total of percentages of his response to the female pictures (D, F, I, K, and M) gave his "response to 'female' picture" score (Table 1, second column). The algebraic subtraction of each subject's male picture total from his female picture total (column two minus column one) gave each subject's relative male-female response measure (Table 1, third column). Using this order of procedure for the table, a positive figure in the third column indicates that the subject had a greater total response to pictures of females than to pictures of males; a negative figure indicates lesser response to pictures of females but greater response to pictures of males.

Results

These male-female response measures clearly discriminate between the subject groups, as is shown in the last column of Table 1. Figure 1 shows this last column graphically. There is

TABLE 1

Pupil Size Increase or Decrease when Comparing Stimuli to Controls Expressed in Percentage Totals

Subject	Total response to 'male' pictures	Total response to 'female' pictures	Relative 'male-female' response score
Heterosexuals			
1	−00.4	+05.9	+06.3
2	−54.5	−22.4	+32.1
3	+12.5	+19.2	+06.7
4	+06.3	+39.0	+32.7
5	−01.5	+23.1	+24.6
Homosexuals			
6	+18.8	+11.2	−07.6
7	−04.6	−38.0	−33.4
8	+18.9	+18.1	−00.8
9	+18.2	−05.6	−23.8
10	+15.8	+21.5	+05.7

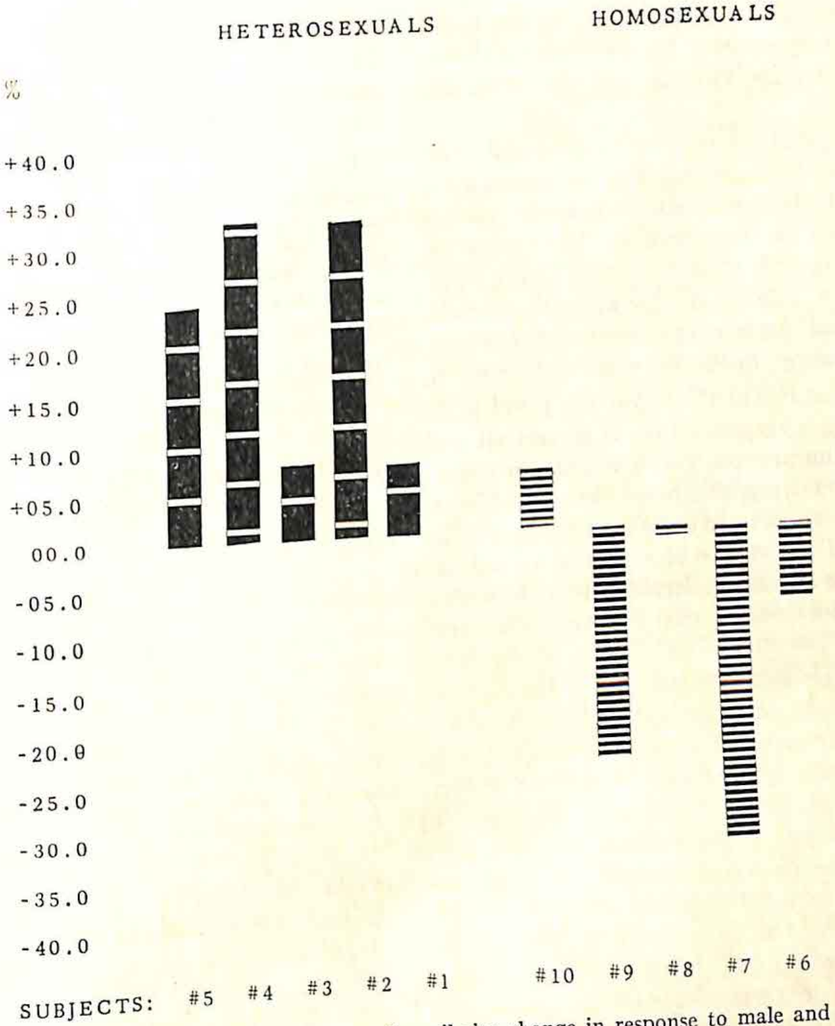


FIG. 1. Differences in percentage of pupil size change in response to male and female pictures. (A positive score shows higher response to pictures of females; a negative score shows higher response to pictures of males.)

no overlap between the groups in that the lowest heterosexual response is +06.3 while the highest homosexual response is no higher than +05.7. All heterosexual males show a larger response to pictures of women than to pictures of men (positive scores). Four of the homosexuals show a larger response to pictures of men (negative scores).

DISCUSSION

Some of the female pictures drew a high-positive response from some of the homosexuals and some of the male pictures drew a high-positive response from some of the heterosexuals. Therefore, response to any single stimulus did not serve to categorize individuals. The total response of a group of subjects to any single stimulus, however, usually served to categorize that stimulus. Total heterosexual response to three of the five female pictures was positive. Total homosexual response to each of the five male pictures was positive. The pictures used in this pilot study were chosen on an a priori basis. The information they have given us and more recent advances in our technique—especially in the matter of brightness matching of pictures—may now permit the formulation of a test battery of pictorial stimuli designed to give a more absolute reflection of a single subject's sex-object interest. It should be emphasized, however, that since *all* subjects in this study saw identical stimuli, the brightness factor could not in any way account for the reported difference between individuals and the resultant groups.

The cooperation of the homosexual subjects, it should be noted, was an unusual relaxation of their customary defense against identification as homosexuals. They were all effectively operating in a normal living environment, in school, at work, with friends. Their sexual preferences were not obvious, and they were ordinarily most reluctant to talk about or reveal them, yet the pupil technique, using a response that is nonverbal and beyond voluntary control, was able to differentiate them from the heterosexual subjects. This is not to say that the pupil response as an index of preference is a predictive substitute for the ultimate criterion of the behavior itself. It does mean that where both preference and behavior are homosexual, even though socially concealed, the pupil response has been shown in this sample to have discriminating power.

Pupil response has already seen application in the area of studies of cognition (Hess & Polt, 1964). In the study of some aspects of personality, compared with projective tests and other instruments and techniques that have been used, this technique appears to us to open up entirely new dimensions.

REFERENCES

HESS, E. H., & POLT, J. M. Pupil size as related to interest value of visual stimuli. *Science,* 1960, **132,** 349–350.

HESS, E. H., & POLT, J. M. Pupil size in relation to mental activity during simple problem-solving. *Science,* 1964, **143,** 1190–1192.

(Early publication received January 9, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 169–179

THE POSTTHERAPY PERIOD[1]

DONALD W. FISKE[2]
University of Chicago

AND

GERALD GOODMAN
University of California, Berkeley

18 mos. after psychotherapy, a variety of tests were readministered to 69 available Ss constituting a biased subgroup from 93 tested previously at the beginning and at the termination of therapy. While the group as a whole showed no reliable movement after therapy, individuals made reliable and correlated changes. Only the Ss' direct report of such change was found to be predictable from several earlier measures. Less favorable movement was found on some measures for the 16 Ss who had additional therapy. While status at termination of therapy appears to be an adequate index of group effects associated with therapy, individuals show systematic trends in the subsequent 18 mos.

The posttherapy period has been used to designate an interval of months or years following the termination of a course of psychotherapy. Studies of this period are relatively few, especially in comparison with the volume of research on the process of psychotherapy and on the changes from pretherapy to posttherapy. The scarcity of research on the posttherapy period may stem in part from an implicit assumption that the individual patient's status at termination provides adequate information on all changes associated with therapy, a belief for which definitive evidence is lacking. Work on this period is also limited because, as in any longitudinal problem, there are marked difficulties in locating subjects and obtaining their cooperation and in determining the influences affecting the subsequent status of the subjects.

The present study is part of a larger program including the analysis of changes associated with psychotherapy (Cartwright, Kirtner, & Fiske, 1963) and the attempt to predict these changes (Fiske, Cartwright, & Kirtner, 1964). The analyses reported in this paper were planned to throw light on three questions: Are there dependable changes in client status during the posttherapy period? What are the directions of any such changes? To what are these changes related?

The few published studies of change in adult outpatients over the posttherapy period display various handicaps. In some studies, the methods of assessment at the follow-up point differed from those at termination of therapy (e.g., Berle, Pinsky, Wolf, & Wolff, 1953; Cowen & Combs, 1950; Luff & Garrod, 1935). Other studies have used relatively few (from 1 to 25) subjects (e.g., Assum & Levy, 1948; Carr, 1949; Haimowitz & Haimowitz, 1952; Jonietz, 1950; Rogers & Dymond, 1954). Compared to these studies, the present investigation has the advantages of a reasonably large *N* (69), a fairly long and consistent interval before retesting (18 months), and a variety of instruments, most of which had already been administered before and at termination of therapy. The purposes of this paper are not only to report a study which in these respects is more satisfactory than earlier ones but also to point out the serious difficulties in investigating the posttherapy period.

PROCEDURE

The basic procedure was simply to test the subjects before therapy, at the termination of therapy, and again 18 months later. The subjects received client-centered therapy at the University of Chicago Counseling Center. While tape-recorded interviews revealed some differences in therapist style, the therapists are a fairly homogeneous group when compared with all therapists with other orientations.

---

[1] We gratefully acknowledge the grants from the Social Science Research Committee of the University of Chicago which made this study possible. The earlier data utilized in this paper were collected with the support of a Ford Foundation Grant to the Counseling Center, University of Chicago.

[2] We are indebted to many people who helped collect or analyze the data: Naomi Berne, Desmond S. Cartwright, Rosalind D. Cartwright, Jean Dames Goodman, William L. Kirtner, William Littlewood, Donald Peterson, Arnold Rachman, Richard Robertson, Anne Maxwell Seiden, Thomas Tyler, Robert Williams, and our colleagues who tested subjects in other cities. We are also indebted to John Shlien and William Schutz for their useful comments on a previous draft of this paper.

The therapists included some who had finished their training recently, as well as some with many years of experience. The median length of experience was 2½ years. The median number of other clients treated by the therapist was 11. For the follow-up group of clients, the median number of interviews was 26.

Eighteen months after the termination point, efforts were made to reach each subject by phone or letter, reminding him of his earlier participation and of the prior notification that there would be a third and final testing. Subjects living within 250 miles of Chicago were tested by a member of the project staff either in the home city or at the University of Chicago. Ten subjects residing at a greater distance were tested by local psychologists who followed prepared instructions.

Methods. The basic data used in the present paper were derived from four client self-report instruments (follow-up questionnaire, self-ratings, *Q* sort, and MMPI), the TAT, and an unstructured interview. Three additional instruments were administered. The Social Attitude Scales and the Sentence Completion were for the most part not used in this study because they had proved unfruitful in the earlier analyses. Another instrument was administered for separate purposes. While full details of these procedures can be found in Cartwright et al. (1963), brief descriptions will be given below.

The follow-up questionnaire consisted of 11 questions. It was patterned after the Client Posttherapy Questionnaire but had appropriate additional items. In addition to such questions as one concerning additional counseling or therapy, four multiple-choice ones were aimed at facets of recent changes:

1. What change has there been in your feelings toward yourself since you terminated? (More discontented, no change, somewhat more contented, much more contented.)
2. What change has there been in the problem(s) that originally brought you to therapy? (Got worse, no change, somewhat resolved, mostly resolved.)
3. How much change in you as a person has occurred since you terminated counseling? (Not changed, changed very little, somewhat changed, changed a good deal.)
4. What change has there been in your feelings toward your life situation? (More discontented, no change, somewhat more contented, more contented.)

For each of these questions, the four alternatives were scored from one to four. A total score was obtained by summing the scores from the four questions.

The self-ratings were made on 14 Behavioral Adequacy (BA) Scales developed for this project. These scales were designed to encompass as completely as possible the variables which had been used in published research on changes associated with psychotherapy. Each scale consisted of 11 to 19 statements, the subject checking the 2 which most closely and next most closely described him. Since the statements were written to represent points along a continuum, a numerical score could readily be obtained from these choices.

The *Q* sort was the same one used in earlier research at the Counseling Center (Rogers & Dymond, 1954). It provided a *Q* adjustment score, the key being based on clinicians' judgments of the items (Dymond, 1953).

The interviews conducted at the time of follow-up testing were not sufficiently controlled and structured. They did yield two usable ratings: the diagnostician-rater's estimate of the subject's level of adjustment at the time of the interview and his estimate of the change in adjustment since the posttherapy testing.

TAT ratings were obtained for 62 of the group, the recordings for the other 7 being too inaudible to provide an adequate transcript. Since these ratings had to be made some time after those for the pre- and the posttestings, and since the diagnostician knew that the protocols were from the follow-up point, the mean levels of these ratings are not emphasized in this report.

Most of the analyses were guided by the factors reported in the first paper (Cartwright et al., 1963). Each factor was represented by the measures previously used, i.e., those with high loadings only on the given factor. Thus the Client Self-Evaluation factor (Factor 1) was represented by the client self-rating on Scale 7 (Choice Process for Courses of Action), the *Q* adjustment score, and MMPI *Si* (Social Introversion)-reversed. The second factor, TAT Adequacy, was represented by the ratings on Scales 1, 4, and 8 (Contentedness with Self, Evaluation of Own Behavior, and Goals-Means Consideration in Courses of Action). Factor 4 was represented by MMPI *Hs* (Hypochondriasis) and *Hy* (Hysteria), the score being reflected so that, as for the other scores, a high score was favorable. Thus a *low* score indicated self-reported somatic symptomatology.

The data in this study are primarily of two kinds: in addition to scores on single measures, composite scores of a special kind were also used, namely, residual gain scores (Manning & Du Bois, 1962) for each of several factors. First, a composite score was obtained for each factor. Each of the scales used to represent the factor was weighted by the reciprocal of its standard deviation so that each scale contributed equally to the composite. These weighted scores were summed to obtain the composite score for each subject. Separate composite scores were obtained from the testing at termination and from the follow-up testing.

The residual gain score provides an estimate of the person's status at the later point with the influence of his prior status partialled out. From the correlation of the scores at the two points (posttherapy and termination), the regression of the second score on the first can be determined. This contribution to the second score is then removed, leaving only that portion which is uncorrelated with the first score.

This statistical procedure yields scores permitting more meaningful comparisons between individuals. It takes the place of the experimental procedure of selecting subjects in such a way that all have the same score at the earlier point in time (in this case, the termination-of-therapy point). The residual gain scores at follow-up are identified below by their initials, RGF. A similar procedure was used in the previous paper (Fiske et al., 1964) to obtain residual gain scores for the posttherapy testing, with the contribution of the pretherapy scores partialled out. These are identified as RG.

In the findings reported below, it may be assumed that the same subjects are used at each point in time for any comparisons between, for example, posttherapy status and follow-up status. Two-tailed tests of significance are used throughout. The .05 level of significance is used unless otherwise indicated.

Subjects. There was a loss of subjects similar to that in most follow-up studies (e.g., Bookbinder, 1962). The reasons for the losses illuminate the potential bias in the experimental sample. Of the 93 subjects used in the previous studies in this program, no reply was received from letters addressed to 4 of them and no effort was made to test 2 who had moved to Europe. Among the 87 (94%) who were located in this country, only 69 (74% of the original 93) were used in this study. Four were dropped because they were still in therapy at the time of their second, "posttherapy" testing; 3 were eliminated because their second testing had occurred more than 3 months after termination of therapy. Also excluded were 4 who were partially tested at the follow-up point but refused to complete some written procedures, 3 who refused to be tested or failed to keep any appointment, and 1 whose materials were lost. Two others were in therapy at the Counseling Center at the 18-month point and the therapist felt that testing at that time would be contrary to the client's best interests. One client was described by her husband as too emotionally sick to be tested. Thus 24 of the original 93 could not be used.

The followed-up cases resembled the parent population in several respects. The group studied had a median age of 26 when they started therapy. Fifty-seven percent were male. The median number of interviews in the follow-up group was 28, little different from that of 26 for the entire group. The range was also very similar.

At the time of the follow-up testing, 54% were married and only 41% were single. In contrast, before therapy, only 39% of the original 93 were married and 54% were single. This shift is more likely a consequence of the high proportion of young adults in the group and of the years between pretherapy and follow-up than a result of therapy or of selection in the follow-up group.

Further therapy. The follow-up questionnaires disclosed that 16 (23%) of the 69 subjects started another course of therapy subsequent to the termination of their therapy at the Counseling Center. (At least 2 of the 24 excluded cases also had further therapy.) Where the original report was vague, letters were written to clarify the pertinent facts. Nine of the 16 were still in treatment at the time of the follow-up testing.

For 2 of the 16 having additional therapy, the therapist was client-centered. Ten of the therapists were psychiatrists, their orientation being described as analytic: Freudian, Jungian, or Adlerian. The profession and orientation of the remaining four therapists could not be determined.

The proportion of cases receiving further therapy corresponds closely with the 25% reported in *The Student Physician as Psychotherapist* (Heine, Aldrich, Draper, Meuser, Tippett, & Trosman, 1962), exceeds the value of 15% noted by Luff and Garrod (1935), but is much less than the figures Frank (1961) gives for several follow-up points.

RESULTS

In the analyses presented below, several different kinds of variables are used in comparing the various groups. Because of this diversity, most of the findings cannot be economically given in tabular form. Therefore a preview seems desirable.

First, two technical matters were considered: The followed-up group was found to be fairly similar in many respects to those not followed-up, although the status of the followed-up group was generally more favorable at the termination of therapy. The measures of changes during the 18 months after the testing at termination were shown to be reliable and, in some instances, to intercorrelate significantly.

Next, the posttherapy and follow-up scores were compared for the group as a whole. No systematic gain or loss was evident.

Finally, the group who had had extra therapy (the X group) in the 18 months following the original course of treatment was subjected to intensive examination. Those who had completed this additional treatment appeared to be similar to those who were still in therapy. In comparison with those having no further therapy (the O group), the total X group had less favorable scores on some measures based on indirect self-report at the follow-up point. Differences between these two groups at pretherapy, during therapy, and at posttherapy were also considered.

Comparison between the Group Followed Up and the Group Excluded

To determine whether the follow-up sample of 69 was representative of the original 93, the group which was followed up was compared with the group which was excluded, in terms of status at posttherapy. Examination of the residual gain scores at termination of therapy (Fiske et al., 1964) showed that the followed-up group was significantly higher (.01 level) on the TAT factor and higher, but not significantly so, on the Client Self-Evaluation factor, the *Hs-Hy* factor, the Interview-Diagnostician score, and the Therapist Evaluation factor. That group was also higher (.01 level) on the therapist's success rating at the end of therapy. The two groups did not differ in length of therapy.

Thus, as of the end of therapy, the relative status of the follow-up group was more favorable than that of the omitted group. The absolute differences were, in general, rather small, so that the variation of scores is not markedly curtailed in the group followed up.

The Dependability of Changes Observed during the Follow-Up Period

While there is sufficient reason to assume that the level or status scores for the follow-up testing point were, in general, internally consistent, there remains the question whether the changes between posttherapy and the later testing are dependable. For some single instruments, there is evidence. Thus the four separate questions on the follow-up questionnaire which were summed to obtain the composite score had a median intercorrelation of .52, from which (by the Spearman-Brown formula) the reliability of the total score can be estimated to be .81. Again, the difference scores between post and follow-up for *Hs* correlated .66 with those for *Hy*.

More telling evidence stems from correlations between different instruments. The change scores in the follow-up period for client Scale 7, *Q* adjustment score, and MMPI *Si*-reversed had a median intercorrelation of .54, permitting the interpretation that the composite score for Factor 1 (Client Self-Evaluation) based on these scales had substantial reliability. Similarly, the residual gain scores for Factors 1 and 4 (RGF-1 and RGF-4) have significant correlations (.25 and .37) with the total score from the follow-up questionnaire. These data permit the interpretation that the change measures from any single self-report instrument and the factor change scores appear to be reliable.

The difference scores (follow-up minus post) for the three TAT ratings had intercorrelations of .49, .60, and .64. The residual gain score derived from these ratings (RGF-2) had a significant correlation with RGF-4 (*Hs-Hy*, reversed). These results allow one to view as reliable the measures of change derived from the TAT ratings.

The importance of this evidence will become apparent later when the lack of systematic group change is reported. While the group as a whole did not change significantly either upward or downward, it would appear that, on particular instruments, some individuals improved while others lost ground during the posttherapy period.

Changes in the Follow-Up Period for the Group as a Whole

Changes in central tendency. Changes during the follow-up period on a number of separate scales were examined. These scales included three components of Factor 1 (Client Self-Evaluation), the two components of the *Hs-Hy* Factor 4, Client Scale 1, Edwards' Social Desirability, and the Sentence Completion measure of Accept Dependency. All but one of these showed no significant change and no clear trend: the exception was the MMPI *Si* scale which showed a significant increase in social introversion. (Subsequent analyses indicated that this one large change was produced by those who had further therapy; see below.)

While similar comparisons of the ratings of the three variables forming the TAT Factor 2 are somewhat questionable, they too provided no consistent or significant evidence of gain or loss.

Changes in group variation. The standard deviations at posttherapy and at follow-up were compared. Again, no systematic differences or trends were found.

Residual gain scores for factors. Of the three intercorrelations between these residual

TABLE 1

CORRELATIONS WITH MEASURES AT FOLLOW-UP

Measure	RGF-1	RGF-2	RGF-4	Questionnaire
Follow-Up				
RGF-1 (Client)		−.03	.14	.25*
RGF-2 (TAT)	−.03		.30*	.17
RGF-4 (*Hs-Hy* reversed)	.14	.30*		.37**
Questionnaire	.25*	.17	.37**	
Posttherapy				
RG-1 (Client)	−.35**	.20	.20	.44**
RG-2 (TAT)	.11	−.07	.09	.25*
RG-3 (Therapist)	.19	−.03	.17	.41**
RG-4 (*Hs-Hy* reversed)	−.09	.09	−.06	.10
RG-ID (Interview)	−.03	−.05	−.12	.21
L (Length)	.02	−.13	.07	.13
C-S (Success rating)	−.10	.01	.17	.30*
Pretherapy				
P-1 (Client)	.29*	.09	−.10	.23
P-2 (TAT)	−.09	.15	.05	.03
P-3 (Therapist)	.05	.18	.17	.24*
P-4 (*Hs-Hy* reversed)	.21	.26*	.27*	.32**
P-ID (Interview)	.12	.28*	.18	.15
Therapist's Experience	.22	−.06	−.08	−.03
Kirtner Typology[a]	−.02	−.16	−.07	.03

Note.—*N*s = 56–69.
[a] The correlations in this row are rhos. All others are Pearson *r*'s.
* $p \leq .05$, two-tailed.
** $p \leq .01$, two-tailed.

gain scores for the follow-up testing, only that between RGF-2 and RGF-4 was significant (see Table 1). However, both RGF-1 and RGF-4 had significant correlations with the follow-up questionnaire.

The three RGF scores had no significant correlations with the residual gain scores (RG) for the posttherapy point, therapist's experience, therapist's success rating, or length of therapy, with one exception (RG-1 versus RGF-1). RGF-1 and RGF-4 each had a significant correlation with its corresponding pretherapy measure. RGF-2 (TAT) is involved in the only other significant correlations with pretherapy factor scores, correlations with *Hs-Hy* reversed and with the score for the Interview-Diagnostician's ratings based on the initial interview (P-ID).

Follow-up questionnaire. In this instrument, each of four direct questions dealing with recent changes as perceived by the patient had comparable response alternatives. The general form was: got worse, no change, somewhat better, much better. Taking these four questions together, the modal response was "somewhat better" and the mean fell on the more favorable side of that category. Thus the typical reply to these direct questions was that the client felt somewhat better at the follow-up point than he felt at the termination of therapy. The direct interpretation of this surprising finding can be discounted, not only because it is not consistent with the absence of mean change on several separate scales administered at posttherapy and at follow-up, but also because it equals the mean found on the questionnaire completed at the end of therapy, an instrument with very similar questions and, for corresponding questions, identical alternatives.

While the mean response has dubious value, the total scores for individuals on these four questions appeared to discriminate between individuals reliably: the separate questions have substantial intercorrelations. Moreover, a large proportion of the correlations between this total score and other variables was significant.

The significant correlations between the total score on the follow-up questionnaire and two of the RGF scores were noted above.

In addition, the questionnaire score had significant relationships with four measures at termination of therapy (see Table 1). While the highest was with Client Self-Evaluation (RG-1), as would be expected, there was also a large correlation with the Therapist's Evaluation (RG-3). Thus the relative favorableness of directly reported gain during the follow-up period was associated with each of several relatively separate measures of gain during therapy. Not given in the table is the correlation of .24 between the follow-up questionnaire and the parallel Client Posttherapy Questionnaire. Although the value was significant (at the .05 level), its modest absolute size suggests that the two questionnaires were not greatly influenced by any common response set, such as social desirability.

The follow-up questionnaire had significant correlations with two pretherapy measures: the *Hs-Hy* factor (P-4) and the Therapist Evaluation Factor (P-3). However, it had no relationship with the typology judgments (Kirtner & Cartwright, 1958) based on the first interview. This typology consists of five ordered descriptions portraying modes of client approach to the resolution of his personal problems.

Thus the client's direct report of relative gain after termination, as given on this questionnaire, has variance in common with two of the other three measures of gain during that period, is predicted by several measures of gain during therapy, and is related to some measures of pretherapy status.

Comparison of Those Having Further Therapy with Others

Considerable emphasis was placed on the minority of the group who reported that they had additional therapy. Why had they sought further treatment? What were the implications of this action for the interpretation of gains as measured at the end of the first course of therapy?

Before comparing the extra therapy group with the others, a possible bias needs to be examined. Nine of the 16 in this group were still in therapy at the time of the follow-up testing. Did this condition affect their scores? When those still in treatment were compared with those who had completed this additional therapy, no differences or clear trends were found on the follow-up questionnaire or the three RGF scores. Those still in treatment had a mean on the *K* scale of the MMPI three raw score points lower than the others. Since the difference did not reach the .10 level of significance, those in treatment cannot be considered to be appreciably less defensive. It was concluded that these two subgroups could be analyzed as a single group.

Comparisons involving the follow-up testing. The X group (the 16 having extra therapy) and the O group (the 53 who did not) did not differ on the composite score obtained from the follow-up questionnaire or on three of its other questions dealing with use of tranquillizers, new activities, and seeking medical attention. The self-descriptions on the 14 BA scales showed only one appreciable difference: the X group was significantly higher (.01 level) on the scale for Evaluation of Own Behavior. The groups were almost identical on *K*.

On the other hand, the O group was significantly higher than the X group on RGF-1 (Client Self-Evaluation) and on RGF-4 (*Hs-Hy* reversed). Detailed inspection of the scores contributing to these measures revealed that the X group had dropped more since the posttesting on the several variables with loadings on each factor, some of the losses being significant at the .05 level. Particularly noteworthy is the increase in *Si* (Social Introversion) which was highly significant for the X group alone and significantly greater than that for the O group. Further evidence that the X group had changed more during the follow-up interval was found in its significantly lower correlation between composite scores for Factor 1 at the post and follow-up points.

On each of the three TAT ratings, the difference between termination and follow-up was more favorable for the X group, but the only significant one was for Scale 4, Evaluation of Own Behavior, the scale which showed a similar significant difference for the self-ratings. However, the groups did not differ dependably on the residual gain score from the TAT ratings (RGF-2).

Another set of independent estimates of

status at follow-up testing and of change since the termination point was provided by the diagnostician who listened to the follow-up interviews. Neither his ratings of adjustment at that point nor his ratings of change since termination inferred from just that interview showed appreciable differences between the X and the O groups.

Comparisons involving the posttherapy testing. The O group was significantly higher than the X group on RG-2 (the residual gain score based on the TAT ratings at termination of therapy). The O group was also higher on RG-3 (the residual gain score based on the therapist's evaluations), but the difference reached only the .10 level. The differences on the other residual gain scores and on the posttherapy questionnaire showed no clear trends.

The therapist's posttherapy rating of success did not differentiate the two groups. On the same questionnaire, the therapists were asked to check applicable statements concerning their feelings about the termination of therapy with each client. For the X group, a significantly higher proportion of therapists checked the statement: "The client has not gone as far as he could, but circumstances beyond his control prevent him from continuing counseling." However, the statement was applied only to minorities in each group (six of the X group and seven of the O group). In the comments in response to an open-ended question on the same instrument, the therapists of the X group tended to report signs of defensiveness and dependency more than did those of the O group, but the difference was not significant.

Comparisons involving other variables. Before the original therapy, the X group was significantly higher (.05 level) on Factor 1, the Client Self-Evaluation factor. They also had higher means on 13 of the 14 BA scales (self-ratings). However, the groups showed trivial differences on MMPI *K*, Dependency (Navran, 1954), and Dominance (Gough, McClosky, & Meehl, 1951). Inspection of "Presenting Problems" as reported in the preliminary interviews revealed great diversity in both groups and no clear trends. There was no difference in the stated duration of these presenting problems. The Accept Dependency ratings based on the Sentence Completion showed no difference for this testing (nor for post or follow-up testing). The groups did not differ on the Kirtner Typology.

Several variables which might be related to the therapy process were also examined: no differences between the groups were found for length of therapy, client age, experience of therapist, sex of therapist or client, and therapist and client being of the same sex versus opposite sexes. Therapy prior to the course of treatment examined in this project was reported by 31% of the X group and 23% of the O group but the trend was not significant. (The percentage of the total group having had prior therapy, 26%, is quite close to the 23% who had further therapy after the posttesting. It should be noted, however, that clients were excluded from the original group if they had had more than six interviews of formal psychotherapy during the preceding 3 years.)

Sixteen of the sample had been studied in another project reporting factors based on the verbal participation of client and therapist in the therapeutic interviews (Rice, Wagstaff, & Butler, 1961; Wagstaff, Rice, & Butler, 1960). Eleven of these 16 were from the O group. The remaining 5 were from the X group; all of these 5 were still in therapy at the time of the follow-up testing. While none of the differences between these small subgroups of 11 and 5 was significant, the marked trends in some instances merit attention as promising leads. As rated from tape recordings of their second interviews, all 5 from our X group were low on Client Factor I (a "Success" factor characterized by the client exploring within himself) and on Client Factor II (a factor describing the client who "talks about matters which include himself, but rarely discusses feelings"), but were high on Client Factor III (a "failure" factor describing the client who "acts as an observer"). Each of these trends was diminished in the next-to-the-last interviews.

Although ratings of therapist participation showed no difference on two therapist factors, none of the five in the X group had therapists who were high on either interview on Therapist Factor III (in which the therapist's responses are "expressed in fresh, connotative language, often with high imagery. . . . The functional level of the responses tends to be that of inner exploration.")

Thus clients seeking further therapy, in at least these instances, tend to commence therapy with a style of participation predictive of poor prognosis but are likely to modify their style during therapy. At neither the beginning nor the end of their initial course of therapy is the participation of their therapist optimal.

Overview. It will be evident that several dozen comparisons were made between the two groups. Hence little confidence can be placed in any one isolated result reaching the .05 level. On the other hand, more confidence can be placed in the findings which form a consistent pattern: e.g., when not only a composite score but also each of its several components show strong and congruent trends.

The major contrast between the X and the O groups appears to be in factors based on self-report measures. Before therapy, the X group evaluated themselves more favorably than did the O group. Between the termination testing point and the follow-up point, the X group's Self-Evaluation factor score moved in the unfavorable direction and their reported somatic symptomatology (*Hs-Hy*) increased. Examination of the data for the separate components of these two factors reveals a strikingly consistent pattern. On each component, the X group had a higher mean than the O group at the posttherapy testing and a lower mean than the O group at the follow-up point. However, between these two points, the X group itself consistently showed unfavorable changes while the O group typically made slight gains toward more favorable scores than its means at posttherapy.

DISCUSSION

Consistent with the trend observed in the few previous studies, this investigation found that the follow-up group as a whole did not show any clear gains or losses during the posttherapy period of a year and a half. This finding fails to support the expectations of some theorists (e.g., Alexander & French, 1946; Rogers, 1951) that gains during psychotherapy provide the basis for continuing improvement after termination. One can, of course, argue that a pattern of continuing gains does occur in some portion of therapeutic cases. However, other cases appear to lose some of the gains they achieve during the therapy period. While status at termination is substantially correlated with status at follow-up testing, the correlation is not of sufficient magnitude to permit dependable predictions of the absolute or relative status of individuals at the later point in time. (Compare the disagreements between evaluations at these two points as reported by Cowen and Combs, 1950. In that study, pooled judgments of "successful," "progress," or "failure" based on the therapy protocols and case summaries agreed with pooled judgments based on the follow-up interview in only 60% of the cases when chance agreement would have been 44%.)

Butler (1960) found 11 types or person-factors in *Q* sorts before therapy and only 4 at follow-up. While the factorial structure at follow-up was not related in any clear way to the first-order factors at pretherapy, it was related to the second-order structure at the earlier point. In the present investigation, however, no clear-cut evidence of increase or decrease in group homogeneity for any continuous dimension was found.

The lack of systematic group change in the period after therapy can be viewed positively. One need not be concerned about the possibility that measures at termination of therapy are temporarily elevated by factors associated with the termination itself. Hence, one can have confidence in such measures as assessments of the status of a group as a whole. The obtained maintenance of status at the follow-up testing must, of course, be interpreted with caution since pertinent data were not available on a number of potential subjects. As noted earlier, the follow-up group was biased toward those who had larger residual gains on several factors at the termination point.

The central tendency of the subjects' responses to the direct questions about improvement in the period after therapy is obviously discounted (cf. Schjelderup, 1955). On the other hand, the follow-up questionnaire emerges as a surprisingly valuable measure of relative change during the posttherapy period. It should be refined. Its four broad alternatives might be replaced by six or seven gradations. Questions comparing changes during this period with those during therapy might prove effective. Several more specific questions might be substituted for each of its general ones. Even in its present crude form, it yields important data on the subject's perception of change in himself, data so comprehensive and broad that they are related to several relatively independent measures of status before and after therapy.

No relationship was found between length of therapy and residual gain scores at the follow-up point. In contrast, significant correlations between length and follow-up status have been reported in one study with a different type of patient and of therapist and with longer average therapy (McNair, Lorr, Young, Roth, & Boyd, 1964).

Further therapy. The study presented here is unusual in its exploration of the characteristics of those who go on to another course of treatment. The obtained results, however, do not fully answer the question of how such people differ from those who do not. One reason is apparent: the follow-up testing was arbitrarily planned for 18 months after termination. Since that time, other subjects may well have reentered treatment. It is likely that the time of such reentry and the decision itself are both influenced by many contingencies associated with the subject's interpersonal environment and other aspects of his life situation.

Heine et al. (1962, pp. 178–179) note that those patients who indicated that they felt better at follow-up than they did at the end of treatment tended not to have had additional treatment in the interim. This finding would seem congruent with the present finding of higher means for the O group than for the X group on the residual gain measures at follow-up based on self-report (RGF-1 and RGF-4). However, no difference was found between groups on the more comparable instrument, the directly worded follow-up questionnaire.

Although evidence concerning the defensiveness and dependency of the X group was vigorously sought, no support was found for the speculation that their return to therapy might stem from their desire to reestablish a dependent relationship. The findings concerning defensiveness were provocative but inconclusive. If the consistently more favorable self-descriptions of the X group before therapy represent defensiveness, it is not of the sort assessed by the *K* scale. Perhaps their prior view of themselves was really more favorable, but some experience after therapy caused them to perceive themselves as less adequate. In some instances, either inability to utilize the initial therapy experience at that time or premature termination appear to have been factors.

Methodological difficulties. The period following therapy needs intensive study. Further knowledge about it is essential to a comprehensive understanding of psychotherapy. The dependable changes occurring during this period require more definitive investigation than was possible in this research. Although the project reported here had several relative advantages, such as a comparatively large *N* and a diversity of measuring instruments used at all three points, it also suffered from many weaknesses common to follow-up studies.

What does this work suggest as the form of an ideal study of the period after therapy? Clients come to therapy with a great range of problems and of personality structures. It seems highly desirable to classify subjects in some way rather than to deal with an assortment of individuals which may not resemble any other collection of cases. One possibility is to restrict the sample to subjects with a given type of difficulty. Another is to classify the subjects post hoc by factor-analytic procedures (cf. Butler, 1960).

Anyone planning to study this period must be prepared to devote the great amount of time and money necessary to locate and evaluate every subject in the original sample. (An illustrious example is provided by Kogan, Hunt, & Bartelme, 1953.) The data reported above provide strong evidence for the conclusion that subjects who are difficult to locate and assess are different from those more available and cooperative. The researcher should be oriented toward a comprehensive longitudinal study of an intact sample.

Again in spite of the expense, a number of independent measures are necessary, measures involving different methods and separate points of view. The present findings are different even for various forms of self-report and self-description. However, in addition to test protocols provided by the subject, the judgments of other people should be sought: what changes have been observed by significant others who know the subject well? What changes have been observed in functioning on the job?

Even more difficult to assess reliably are changes in the subject's environment. Has the subject's family constellation changed in composition? Have there been changes in the conditions associated with his work? These questions and the preceding ones could be answered with confidence only on the basis of a careful social investigation.

The posttherapy period is difficult to study because changes are typically small, and many factors appear to be involved. In addition to approaching the problem through the use of homogeneous groups, the use of composite scores seems necessary. Although changes on any single rating or test variable appear to have low reliability, the results for combined scores tend to be more clear-cut and consistent. Each composite score should have substantial intercorrelations among its components. It also should be made conceptually significant by having it represent a distinctive aspect of personality as assessed by a particular method or a separate vantage point for viewing the subject.

SUMMARY

This investigation involved a large number of statistical analyses, most of which are reported above. With such a quantity, one must expect to obtain some "significant" results where no true difference or relationship exists. On the other hand, congruent and interrelated results deserve emphasis and such findings are summarized below.

1. Cases were lost to this follow-up study for diverse reasons, the available cases constituting a biased subgroup.

2. Measures of individual moment in the posttherapy period, especially composite scores, were typically reliable.

3. The total group showed no significant movement during this period.

4. Measures of posttherapy movement show some covariation over individuals. Self-reported answers to a set of direct questions were correlated with movement measured by different instruments.

5. While scores indicating either status before therapy or adjusted gain during therapy did not predict posttherapy movement as assessed by residual gain scores, some of these scores did predict the subjects' direct report of such movement.

6. Those who had additional therapy during this period were lower on the two movement measures from indirect self-report but were comparable to the other subjects on scores from the direct self-report instrument and from TAT ratings.

7. Some trends suggest that the group with additional therapy evaluated themselves more favorably at the pretherapy testing but in their early therapy interviews were less prepared to utilize the therapy experience. This group may provide the most striking examples of the complex longitudinal process occurring during the therapy and posttherapy life-segment.

REFERENCES

ALEXANDER, F., & FRENCH, T. M. *Psychoanalytic therapy: Principles and application.* New York: Ronald Press, 1946.

ASSUM, A. L., & LEVY, S. J. Analysis of a nondirective case with follow-up interview. *Journal of Abnormal and Social Psychology,* 1948, **43,** 78-89.

BERLE, BEATRICE B., PINSKY, RUTH H., WOLF, S., & WOLFF, H. G. Appraisal of the results of treatment of stress disorders. *Research Publications Association for Research in Nervous and Mental Disease,* 1953, **31,** 167-177.

BOOKBINDER, L. J. Follow-up versus discharge status of psychiatric inpatients. *Journal of Clinical Psychology,* 1962, **18,** 501-503.

BUTLER, J. M. Self-concept change in psychotherapy. University of Chicago: *Counseling Center Discussion Papers,* 1960, **6,** No. 13.

CARR, A. C. An evaluation of nine nondirective psychotherapy cases by means of the Rorschach. *Journal of Consulting Psychology,* 1949, **13,** 196-205.

CARTWRIGHT, D. S., KIRTNER, W. L., & FISKE, D. W. Method factors in changes associated with psychotherapy. *Journal of Abnormal and Social Psychology,* 1963, **66,** 164-175.

COWEN, E. L., & COMBS, A. W. Follow-up study of 32 cases treated by nondirective psychotherapy. *Journal of Abnormal and Social Psychology,* 1950, **45**, 232–258.

DYMOND, ROSALIND F. An adjustment score for *Q* sorts. *Journal of Consulting Psychology,* 1953, **17**, 339–342.

FISKE, D. W., CARTWRIGHT, D. S., & KIRTNER, W. L. Are psychotherapeutic changes predictable? *Journal of Abnormal and Social Psychology,* 1964, **69**, 418–426.

FRANK, J. D. *Persuasion and healing: A comparative study of psychotherapy.* Baltimore: Johns Hopkins Press, 1961.

GOUGH, H. G., MCCLOSKY, H., & MEEHL, P. E. A personality scale for dominance. *Journal of Abnormal and Social Psychology,* 1951, **46**, 360–366.

HAIMOWITZ, NATALIE K., & HAIMOWITZ, M. L. Personality changes in client-centered therapy. In W. Wolff & J. A. Precker (Eds.), *Success in psychotherapy.* New York: Grune & Stratton, 1952. Pp. 63–93.

HEINE, R. W. (Ed.), ALDRICH, C. K., DRAPER, E., MEUSER, MARY, TIPPETT, JEAN, & TROSMAN, H. *The student physician as psychotherapist.* Chicago: Univer. Chicago Press, 1962.

JONIETZ, ALICE K. A study of the phenomenological changes in perception after psychotherapy as exhibited in the content of Rorschach percepts. Unpublished doctoral dissertation, University of Chicago, 1950.

KIRTNER, W. L., & CARTWRIGHT, D. S. Success and failure in client-centered therapy as a function of initial in-therapy behavior. *Journal of Consulting Psychology,* 1958, **22**, 329–333.

KOGAN, L. S., HUNT, J. MCV., & BARTELME, PHYLLIS. *A follow-up study of the results of social casework.* New York: Family Service Association of America, 1953.

LUFF, MARY C., & GARROD, MARJORIE. The after-results of psychotherapy in 500 adult cases. *British Medical Journal,* 1935 (July) **13**, 54–59.

MCNAIR, D. M., LORR, M., YOUNG, H. H., ROTH, I., & BOYD, R. W. A three-year follow-up of psychotherapy patients. *Journal of Clinical Psychology,* 1964, **20**, 258–264.

MANNING, W. H., & DU BOIS, P. H. Correlational methods in research on human learning. *Perceptual and Motor Skills,* 1962, **15**(2), 287–321.

NAVRAN, L. A rationally derived MMPI scale to measure dependence. *Journal of Consulting Psychology,* 1954, **18**, 192.

RICE, LAURA N., WAGSTAFF, ALICE K., & BUTLER, J. M. Some relationships between therapist's style of participation and measures of case outcome. University of Chicago: *Counseling Center Discussion Papers,* 1961, **7**, No. 5.

ROGERS, C. R. *Client-centered therapy: Its current practice, implications, and theory.* Boston: Houghton Mifflin, 1951.

ROGERS, C. R., & DYMOND, R. F. (Eds.) *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

SCHJELDERUP, H. Lasting effects of psychoanalytic treatment. *Psychiatry,* 1955, **18**, 109–133.

WAGSTAFF, ALICE K., RICE, LAURA N., & BUTLER, J. M. Factors of client verbal participation. University of Chicago: *Counseling Center Discussion Papers,* 1960, **6**, No. 9.

(Received February 17, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 180–187

CONDITIONING OF AFFECTIVE SELF-REFERENCES AS A FUNCTION OF THE DISCRIMINATIVE CHARACTERISTICS OF EXPERIMENTER INTERVENTION

MICHAEL MERBAUM[1] AND EUGENE A. SOUTHWELL[2]
University of Chicago

This investigation tested the effects of 2 different E interventions on the conditioning of affective verbal behavior. The interview was divided into 3 10-minute periods: operant level, acquisition, and extinction. During acquisition E either echoed S's affective self-reference or paraphrased its content. In a control treatment, E arbitrarily echoed or paraphrased a comparable number of nonaffective comments. Following extinction Ss listed all affective words they remembered E saying. Results indicated that the paraphrase was significantly most effective in influencing Ss' verbal behavior. The paraphrase group also recalled a significantly greater number of affective words expressed by E. The results were interpreted as indicating the dominant role of discriminative stimuli in verbal conditioning performance.

In the past few years a growing number of verbal conditioning studies have focused on an experimental analysis and manipulation of the quasi-clinical interview (Adams & Hoffman, 1960; Merbaum, 1963; Moos, 1963; Rogers, 1960; Salzinger, 1960; Waskow, 1962). In this experimental arrangement, the experimenter is "programmed" to respond systematically to a preselected class of verbal behavior while preserving the spontaneous character of the interview transaction. Because this experimental setting is designed to be more or less naturalistic and since the verbal operants influenced have frequently had a personal affective quality, extensions of findings have often been made to psychotherapy and other social influence situations (Krasner, 1962; Marmor, 1962).

In the majority of the verbal conditioning interview studies, the programmed interventions have consisted of simply reinforcing stimuli such as "mmm-hum," "uhhuh," "I see," "good," etc., often in conjunction with nonverbal gestures of approval (Adams & Hoffman, 1960; Moos, 1963; Rogers, 1960; Salzinger, 1960). While reported results suggest that simple selective attention is capable of influencing the emission rate of a variety of verbal operant classes, it is unlikely that the most productive real life conversations are carried out under conditions where such minimal audience participation is required. As Skinner (1957) aptly comments, "Verbal behavior would be pointless if the listener did nothing more than reinforce the speaker for emitting it [p. 151]." With few exceptions (Merbaum, 1963; Waskow, 1962) verbal conditioning research has not explored the effects of more complex content interventions by the experimenter in the manipulation of the subject's verbal behavior.

The purpose of this study is to investigate the effects of two forms of experimenter content interventions (a paraphrase and echoic response) in the conditioning of self-referred affective verbal behavior. It is predicted that the experimenter's introduction of slightly original stimuli (paraphrase condition) will facilitate a greater number and variety of self-referred affect word expression than either repetitive verbal stimuli (echoic condition) or a nonaffect control condition in which affect material is completely ignored. This prediction is based on the assumption that experimenter interventions in the verbal conditioning experiment can be conceived of mainly as discriminative stimuli which set the occasion for the emission of previously reinforced verbal behavior. Thus, the paraphrase, by introducing variations in affective content, should provide clear-cut external cues for the evocation of related verbal responses by the subject. The echoic response, by providing a less discriminative external cue, should restrict the range of the subject's associative responding in comparison to the paraphrase treatment.

---

[1] Now at Bowling Green State University.

[2] Now at University of Indiana, Northwest Campus.

METHOD

Subjects

Thirty junior medical students served as subjects. All were, at the time of their participation in the study, engaged in a 1-month psychiatric inpatient rotation as part of their medical training. Each rotation group was composed of six students. In all, five student groups were interviewed over a period of 6 months. Subjects consented to be interviewed about their inpatient experiences during the second week of the psychiatric rotation training term.

Interview Procedure

Each subject was interviewed for 30 minutes. The instructions for this interview were designed to emphasize a discussion of "feelings" and emotional experiences. It was hoped that these instructions would initially place a high priority on emotional material in order to obtain basal levels of affective verbal behavior under equally optimal conditions for all groups. The instructions for the interview were as follows:

> I would like you to describe your feelings about the inpatient ward. How you fit in, your relations with patients and staff and any feelings you might have about the structure and goals of the ward. You can discuss personal feelings about the patients you have most contact with, how you feel about the group meetings or any other personal experiences you have found interesting. Perhaps we might start out with any general feelings you have about the unit.

In order to investigate the effects of verbal interventions of the interviewer on the affective verbal behavior of subjects, each interview was divided into three 10-minute periods: operant level, acquisition, and extinction. In the operant level period, the base rate of affective response emission was established. Although no critical experimenter interventions were administered during this period, the experimenter intervened on occasion to ask questions. During all three periods, questions were asked only if subject remained silent for more than 10 seconds. An attempt was made to ask the same questions at the same frequency over all interview periods for all treatment groups. In the acquisition phase, the experimenter responded with either a paraphrase or echoic comment following the emission of all self-referred affective words or phrases. During extinction, the experimenter withheld all systematic responding while continuing to present general acknowledgments (smiles, nods, etc.) to maintain the conversational rapport.

Dependent Variables

The response class of *self-referred affective words* was defined as any word directly referring to or identifying a personally relevant emotional experience or state. The personalized character of the affect word was usually established if it was preceded by an "I" or "We" pronoun. If, for example, the subject remarked, "I am afraid, anxious, happy, and depressed, pleased, upset, etc.," the words contained within this affective self-reference unit were individually scored. Thus, for each subject this score consisted of the frequency of self-referred affect words during each interview period.

A second scoring category consisted of the number of *different affect words* each subject used during each separate interview period. When the subject repeated the same word in the same interview period, only one score was recorded for that word. However, if the same word appeared in a different interview period, it was counted as an affective word in that separate period.

The frequency of *affective self-reference statements* was also computed for each subject for each interview period. This response class is identical to the one described by Salzinger (1960) as "any statement describing or evaluating the state (other than intellectual or physical) of the (subject) by himself." This response class included only those affect statements which began with the pronouns "I" or "We."

Finally, a frequency count was made of the total *nonaffective "I" and "We" statements* appearing within each separate interview period. This response variable provided a rough estimation of the comparative self-reference fluency of the respective treatment groups and indicated whether the orientation of subject's discussion centered more on affective or nonaffective topics.

Thus, there were four dependent variables for each separate interview period: (*a*) total number of self-referred affective words; (*b*) number of different self-referred affect words; (*c*) total number of affective self-reference statements; (*d*) total number of nonaffective self-reference statements.

Experimental Treatments

Thirty subjects were randomly assigned to three treatment groups. Each group was composed of 10 subjects.

Subjects in the *paraphrase group* received a paraphrase of the affective word contained in their affective self-reference statement immediately following its emission. For example, if subject remarked, "I feel angry," the experimenter might comment, "You feel irritated" or "You feel mad." Although the paraphrase intervention was presented arbitrarily on the basis of what "felt right" at the moment, an effort was made to keep the paraphrase word relatively equivalent to the subject's affective content. Occasionally the experimenter intervened with a general word covering a class of emotional statements such as, "You feel uncomfortable or pleased," depending upon the negative or positive character of the subject's statement. The experimenter tried to use a variety of emotional words without duplicating any of his own paraphrases during the acquisition phase of the interview. This was not always successful, however, and duplications did occasionally result. The experimenter reinforced only explicitly stated affective words. Thus, when a feeling was implied in action or tone but not verbally expressed as such, the experimenter did not attempt to paraphrase. The experimenter always attempted to capture the verbalized feeling predominant at the moment by providing an appropriate paraphrase. Although experimenter attempted to intervene immediately following an affective word, there was, at times, a delay due to the continuous nature of the subject's commentary. Also, in some instances, the experimenter could not intervene at all if the subject went off into a completely different tangent of thought. Interventions were thus administered on a noncontinuous reinforcement schedule.

Subjects in the *echoic group* were given a repetition of the affective word contained in their affective self-reference statements. Thus, if the subject remarked, "I feel angry" the experimenter would comment, "You feel angry," in an understanding and encouraging tone of voice. As with the paraphrase condition, the experimenter could not always intervene immediately following the affective word. Thus, in this treatment condition, also, interventions were usually administered on a noncontinuous reinforcement schedule.

The experimental arrangements for the *control group* paralleled those of the repetition and paraphrase groups except that in the acquisition period a relatively standard number of interventions were administered to arbitrarily selected nonaffective I-We responses. For example, if the subject stated, "I was curious or interested," the experimenter might respond, "You are a curious person" or "You seem interested in finding new things to do." In this treatment condition, no attention was ever given to any affective material. In addition, the number of experimenter interventions during the acquisition period was of the same magnitude as in the other two treatment conditions, the intent being to equalize the experimenter's activity level in each of the three treatment groups.

Awareness Interview

Immediately following the conclusion of the interview, each subject was handed a sheet of paper containing the following four questions concerning the interview:

1. The purpose of the interview was ______.
2. My evidence for this is ______.

On a second sheet of paper were a set of questions presented to the subject after the first two had been completed:

3. List all the feeling words you expressed about yourself during the interview ______. For example, the words "happy" and "sad" would be feeling words.
4. List all the feeling words that you remember the interviewer said during the interview ______.

After the subject read the first two questions, the experimenter asked if there were any questions. He then left the room for a period of 10 minutes. On his return, the experimenter handed the subject the final two questions and again left the room.

RESULTS

Preliminary Analysis of Interview Data

Before the data pertaining to the hypotheses relative to conditioning as a function of differential experimenter intervention could be analyzed, it was necessary to assess whether the experimenter's attempt to ask "interview continuing" questions at equal frequencies throughout all three 10-minute periods of the interview was successful. It was also necessary to test whether experimenter interventions occurred at the same frequency in all groups. In addition, reliability of the measurement of the dependent variables was also evaluated.

The results of t tests carried out on all pairs of the mean number of "interview continuing" questions asked by the experimenter in each interview period for each treatment group were statistically insignificant (all p's $> .20$). Thus, any possible effects of this type of experimenter behavior on the verbal behavior of the subject's was equated over all interview periods and treatment conditions thereby allowing one to place more confidence in the effects of the independent variables. In addition, the frequency of the experimenter intervention (echoic, paraphrase, and nonaffective) did not vary significantly between the two experimental groups or between each of the experimental groups and the control group (all t tests yielded p values $> .20$). Thus, the frequency of experimenter intervention was equivalent in each of the three groups.

To assess the reliability of the coding of both experimenter and subject interview behavior, a second judge independently coded three randomly selected interviews from each of the three treatment groups. In terms of both frequency of experimenter intervention as well as the four dependent variables, interrater reliability correlations ranged from .79 to .96, indicating that the experimenter's coding of interviews was not a significant contributor to any differential effects of experimental treatment.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF BOTH AFFECTIVE AND NONAFFECTIVE SELF-REFERENCES FOR ALL EXPERIMENTAL GROUPS IN ALL INTERVIEW PERIODS

		Affective self-references			Nonaffective self-references		
		Operant level	Acquisition	Extinction	Operant level	Acquisition	Extinction
Paraphrase	*M*	8.20	17.60	9.90	24.60	20.40	20.70
	SD	2.20	5.02	3.50	4.91	5.50	5.38
Echoic	*M*	8.50	11.60	5.70	23.40	18.90	22.90
	SD	3.98	5.04	2.71	5.35	5.71	4.28
Control	*M*	8.50	5.70	4.50	24.60	29.20	25.80
		4.49	2.22	2.72	7.96	7.54	7.40

Conditioning of Affective Self-References

The mean frequencies of affective self-references during the operant level, acquisition, and extinction periods for the different treatment groups are presented graphically in Figure 1. The means and standard deviations of these data are presented in Table 1. These data were submitted to two separate Type I analyses of variance (Lindquist, 1953), one on the operant level-acquisition periods and the other on the acquisition-extinction periods. Analysis of the operant level-acquisition data yielded significant main effects for experimental treatments ($F = 7.41$ for 2 and 27 *df*; $p < .005$) and for interview periods ($F = 16.68$ for 1 and 27 *df*; $p < .001$) and a significant treatments by periods interaction ($F = 19.83$ for 2 and 27 *df*; $p < .001$). Comparisons of individual pairs of means by t test reveal that only the paraphrase group significantly increased in rate of affective self-reference emission from operant to acquisition periods ($t = 3.06$ for 27 *df*;[3] $p < .01$). None of the other within-group comparisons were significant (all p's $> .10$). Between-group comparisons show that the paraphrase group emitted a significantly greater number of affective self-references during acquisition than did the echoic group ($t = 3.28$ for 18 *df*; $p < .005$). Also, the paraphrase group differed significantly from the control group ($t = 6.51$ for 18 *df*; $p < .001$) as did the echoic group ($t = 3.22$ for 18 *df*; $p < .005$). Thus, the paraphrase treatment was significantly more effective in increasing the number of affective self-references during acquisition than either the echoic or the nonaffect (control) conditions. Furthermore, the paraphrase intervention was the only treatment to demonstrate a statistically significant increase in affective self-references over its initial operant level. Although the echoic group also increased in response rate during acquisition, this change was not statistically significant. The nonaffect control group, on the other hand, declined in affective self-responding in each subsequent interview period.

Analysis of the acquisition-extinction data produced significant main effects for experimental treatments ($F = 20.78$ for 2 and 27 *df*; $p < .001$) and for interview periods ($F = 44.15$ for 1 and 27 *df*; $p < .001$) and a significant treatments by periods interaction ($F = 6.81$ for 2 and 27 *df*; $p < .001$). Within-

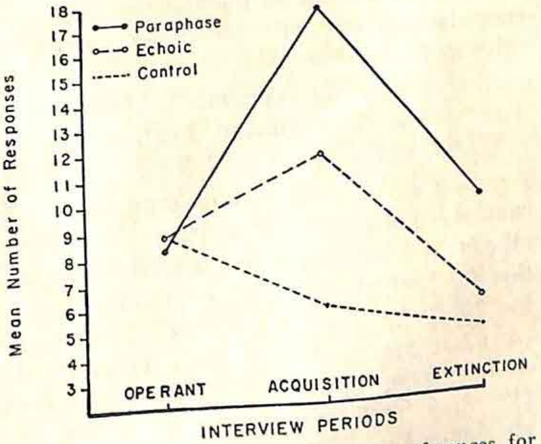


FIG. 1. Mean number of affective self-references for each interview period.

---

[3] The error terms used in within group t tests were derived from the Type I analysis of variance since they presumably are more stable error estimates.

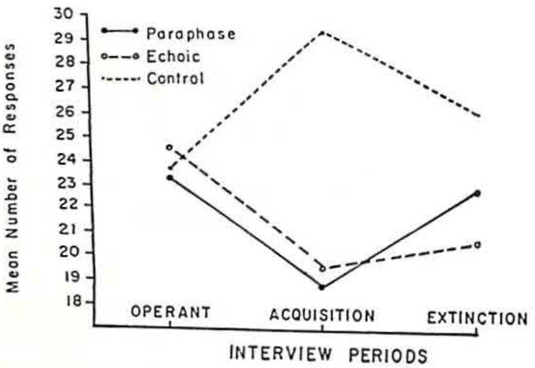


FIG. 2. Mean number of nonaffective self-references for each interview period.

group comparisons showed a significant decline in emission of affective self-references from acquisition to extinction for both the paraphrase group ($t = 6.00$ for 27 df; $p < .001$) and the echoic group ($t = 4.59$ for 27 df; $p < .001$). The control group changes were not significant. Within-group comparisons between operant level and extinction yielded nonsignificant t values (all p's $> .20$). Thus, although the three groups maintained the same relationship to one another as was found in acquisition, withdrawal of selective attention in the paraphrase and echoic groups resulted in a decrease in affective self-reference rate for both these groups.

Conditioning of Nonaffective Self-References

The nonaffective self-reference data were analyzed using a Type I analysis of variance (Lindquist, 1953) in a manner similar to the analysis carried out on the affective self-reference data. These data are presented graphically in Figure 2.

Analysis of the operant level-acquisition data yielded a significant treatments-by-periods interaction ($F = 7.57$ for 2 and 27 df; $p < .005$). All main effects were nonsignificant. Within-group comparisons show a significant decline from operant level to acquisition for both the paraphrase group ($t = 2.23$ for 27 df; $p < .025$) and the echoic group ($t = 2.39$ for 27 df; $p < .025$). However, the control group showed a significant increase ($t = 2.45$ for 27 df; $p < .025$). There were no significant differences between groups at the operant level although during acquisition there were significant differences between the paraphrase group and control ($t = 3.15$ for 18 df; $p < .005$) and the echoic group and control ($t = 3.68$ for 18 df; $p < .001$).

Analysis of the acquisition-extinction data resulted in a significant main effect for experimental treatments ($F = 5.54$ for 2 and 27 df; $p < .01$) as well as a significant treatments by periods interaction ($F = 3.68$ for 2 and 27 df; $p < .05$). Within-group comparisons yielded nonsignificant t values (all p's $> .10$) while between-group comparisons show a significant difference between the paraphrase and control group during extinction ($t = 2.00$ for 18 df; $p < .05$).

These analyses lead to the conclusion that conditioning occurred for the nonaffective control group while the two experimental groups significantly decreased in their use of nonaffective self-references.

Number and Variety of Emotional Words

Similar analyses of variance were carried out on the number of affective self-referred words. These results were almost identical to the affective self-reference findings already reported, indicating a high correlation between the frequencies of emission of affective self-reference and affective words. Analysis of the variety of affective word data also yielded similar results. In order to test whether the large differences in variety of affective word usage was independent of the gross number of affective words used by the subjects, a covariance analysis was carried out on the variety of affective words data while controlling for the total number of affective words. No significant findings were observed suggesting that variety of affective words was dependent upon the number of affective words expressed.

Awareness

None of the subjects in either of the three experimental groups were able to verbalize awareness of any relationship between their verbal behavior (either affective or nonaffective) and the experimenter's verbal interventions. Despite this apparent lack of awareness, 80% of the subjects in the paraphrase group indicated that the expression of feelings or emotional reactions constituted the main purpose of the interview. This is in contrast to 50% of the echoic subjects and only 20% of the nonaffect subjects who conceived the purposes of the interview in similar terms. Relative to these findings, it is important to remember that the feeling set was introduced to *all* subjects at the beginning of the experiment. Thus, it would appear that the nature of the experimenter's verbal interventions (paraphrase or echoic) was highly influential in maintaining this initial set. In comparing the response rates of the three experimental groups, the paraphrase group was the highest in affective responding as well as in verbalizing an awareness of the affective set more readily than the other two groups. It is of interest to note that despite the initial feeling set given the control subjects, this group shifted to a nonaffective intent for the interview.

The final two questions in the awareness inventory focused on (*a*) the subject's memory of words he himself emitted during the interview and (*b*) the number of affective words that the subject remembered the experimenter using during the interview. The mean number of self-expressed affective words remembered was 4.5 for the paraphrase group, 4.6 for the echoic group, and 2.2 for the nonaffect group. Comparisons by *t* test between the paraphrase and nonaffect control and the echoic and nonaffect control groups relative to these data show that both experimental groups differed significantly from the nonaffect control group (paraphrase-control: $t = 3.85$ for 18 *df*; $p < .005$); (echoic-control: $t = 2.72$ for 18 *df*; $p < .01$) but not from one another. Mean number of words attributed to the experimenter were 3.0 for the paraphrase group, 1.1 for the echoic, and .01 for the nonaffect group. Comparisons between the paraphrase and echoic group were statistically significant ($t = 2.84$ for 18 *df*; $p < .01$). Matching the words remembered and the actual words expressed in the interviews by both the experimenter and the subjects indicate that the subject's memory for words was generally accurate. Thus, the words which the subjects attributed to themselves and the experimenter, were to a large extent, the actual words used in the interview.

DISCUSSION

The prediction that the paraphrase treatment would result in the emission of a greater mean frequency of affective words and self-references than in the echoic or nonaffect control groups was supported. In the echoic group, nevertheless, there was a moderate increase in affective responding from operant to acquisition but the difference was not statistically significant. In contrast, the nonaffect control group gradually declined in affective word emission over the three interview periods.

An additional prediction that the variety of affective self-referred words would be greater in the paraphrase group was not confirmed. Although the paraphrase group emitted a significantly greater variety of words than either of the other two groups, this finding proved to be a function of the number of words emitted by this group rather than due to the experimental treatment as such.

Focusing on the functional characteristics of the experimenter's verbal interventions the results suggest that one explanation for the greater effectiveness of the paraphrase stimuli in increasing the subject's emission of affect statements lies in the more discriminable characteristics of the paraphrase stimuli. This explanation is further supported by the awareness data which reveal that the paraphrase group recalled a significantly greater number of the experimenter's verbal responses than did the echoic group even though the experimenter intervened at approximately the same frequency with each treatment group. In evaluating this difference, it is important to note that an extinction period preceded the awareness inquiry, thereby probably increasing the proportion of losses in retention of this material. Moreover, since both the paraphrase and echoic subjects remembered an almost equal number of personal self-referred affective words, the possibility that the intervening extinction period was deleterious to the echoic group seems unlikely. A more reasonable interpretation is that the higher response rates in the paraphrase group are at least partially attributable to the subject's attention to the experimenter's comments. This result would seem compatible with Rogers' (1942) early observation that the reflection of feeling is considerably more effective than reflection of content in bringing about personality change within the psychotherapeutic interaction. In the present study, reflection of feeling is comparable to the paraphrase treatment while reflection of content would be similar to the echoic treatment. The implication is, therefore, that in Rogerian psychotherapy, the reflection of feeling, as a therapeutic strategy, sets the occasion for high rates of affective verbal responding.

Extending this analysis further, the structure of the paraphrase intervention also added an additional aspect to the conversation, a different emotional word, thus providing discriminative stimuli for the emission of related verbal associations. It was found, for example, that following the presentation of a paraphrase remark, there appeared a chaining of affective verbal associations which tended to persist until the content of the topic being discussed had become momentarily exhausted. In the echoic group, such chaining phenomena rarely occurred with the same frequency. For these subjects the continuance of emotional responding appeared to be due more to the subject's private involvement in his own commentary than to his responsiveness to external intervention.

The question arises as to whether the echoic response of the experimenter was aversive, thereby producing an inhibition of affective responding. When the rates for total self-reference emission, both the affective and nonaffective combined, are compared no significant differences were found either between-groups or between-interview periods (all p's $> .10$). Thus, all groups were responding at a comparable level of self-reference frequency during each interview period though the emphasis on affect or nonaffect material in the three treatment groups varied depending upon the experimenter's behavior.

The performance of the nonaffect control group indicates that conditioning occurred for nonaffective self-references. This is further supported by the awareness data for the nonaffect control which indicates that virtually no affective words were attributed to the experimenter and, in comparison to other groups, significantly less self-referred affective words were listed. It thus seems apparent that counter to the initial instructional set, the experimenter was able to successfully lead the subject away from the expression of affective material.

The sharp decline by both experimental groups in affective responding during extinction to a rate roughly equivalent to the initial operant level makes it clear that no new learning or even moderate alterations in verbal habit structure occurred. Further, it is evident that rate of response during acquisition had little influence on resistance to extinction. Consequently, the verbal conditioning effects observed during acquisition seem quite dependent upon the momentary audience variables which control verbal behavior in a dyadic interaction. Since the verbal response to be influenced has had a long history of previous conditioning, it would seem that the task of the subject in the verbal conditioning experiment lies more in the discrimination and selection among a series of prior learned verbal responses than in the acquisition of additional verbal responses. Further, not only has the verbal operant been previously acquired, but the conditions under which the audience exerts control over its emission have been established as well. Consequently, the frequently reported transiency of verbal conditioning phenomena in studies showing sharp decrements in response frequency during extinction is consonant with the interpretation that once the audience variables take on a different character (verbal discriminative stimuli as well as positive reinforcement absent) the stimulus complex of the situation has been crucially altered. Viewed in this way, the subjects reaction to extinction conditions by reduced operant frequency can be regarded as an appropriate discriminative response to the modified social verbal requirements presented by the experimenter. Thus, if discrimination is the major psychological principle involved, it would be predicted that the subjects' verbal repertoire should return to its initial response rate during extinction. The data presented here support such an explanation.

On the whole, research in verbal conditioning seems to have neglected the important role of the discriminative stimulus in the control of verbal behavior. Some authors, notably Matarazzo, Saslow, & Paresis (1960), and Skinner (1957), have pointed to this interpretation. Otherwise, emphasis has usually been placed on the reinforcing or the implied rewarding characteristics of the experimenter's responses. The results of the present study strongly suggest that if mere audience acknowledgment of a response were sufficient, then no difference between the echoic and paraphrase treatments should have resulted. If, however, positive acknowledgment as generalized reinforcement is held constant, then the discriminative properties of the content of various stimuli should play a major role, as they did, in the selectiveness of response emission. Also, since the verbal behavior evoked in this interview situation has had a long reinforcement history, what is most critical for further research would seem to be the isolation of those variables which set the occasion for the discriminated operants to appear.

REFERENCES

Adams, J. S., & Hoffman, B. The frequency of self-reference statements as a function of generalized reinforcement. *Journal of Abnormal and Social Psychology*, 1960, **60**, 384–389.

Krasner, L. The therapist as a social reinforcement machine. In H. Strupp & L. Lubrosky (Eds.), *Research in psychotherapy*. American Psychological Association, 1962.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

Marmor, J. Psychoanalytic therapy as an educational process. In, *Science and Psychoanalysis*, Vol. 5. New York: Grune & Stratton, 1962.

Matarazzo, J., Saslow, G., & Paresis, E. N. Verbal conditioning of two response classes: Some methodological considerations. *Journal of Abnormal and Social Psychology*, 1960, **61**, 190–206.

Merbaum, M. The conditioning of affective self-reference by three classes of generalized reinforcers. *Journal of Personality*, 1963, **31**, 179–191.

Moos, R. H. The retention and generalization of operant conditioning effects in an interview situation. *Journal of Abnormal and Social Psychology*, 1963, **66**, 52–58.

Rogers, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.

Rogers, J. M. Operant conditioning in a quasi-therapy setting. *Journal of Abnormal and Social Psychology*, 1960, **60**, 247–252.

Salzinger, K. Reinforcement of verbal affect responses of normal subjects during the interview. *Journal of Abnormal and Social Psychology*, 1960, **60**, 127–130.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Waskow, I. E. Reinforcement in a therapy-like situation through selective responding to feelings or content. *Journal of Consulting Psychology*, 1962, **26**, 11–19.

(Received February 21, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 188–193

IMPACT OF PSYCHOTHERAPY ON THE PRODUCTIVITY OF PSYCHOLOGISTS

LAUREN G. WISPÉ [1]

Office of Manpower, Automation and Training,
United States Department of Labor, Washington, D. C.

AND

MORRIS B. PARLOFF

National Institute of Mental Health

To investigate (a) the effects of psychotherapy on the scientific publications of American psychologists and (b) the extent to which this group had received psychotherapy, questionnaires were sent to all APA members receiving PhDs 1945–1951 to obtain information about their therapy. A number of publications was obtained from the *Psychological Abstracts*. There were no significant differences in number of publications between therapy and matched nontherapy control samples. Of 966 male psychologists responding, $\frac{1}{3}$ had received psychotherapy, with clinicians disproportionately represented. Of this $\frac{1}{3}$, 13% had 3 or more therapeutic experiences. There was a high positive correlation between duration of therapy and therapy satisfaction.

The purposes of the present study were to assess the impact of psychotherapy upon scientific writing, and to investigate the extent and kinds of psychotherapy received by American psychologists with various professional affiliations.

Psychotherapeutic outcome has been judged in different ways: alleviation of symptoms, improved peer and familial relations, veridical perception of reality, and the ability to work and contribute to society. These are not mutually exclusive categories, and the present study emphasized the last one.

Scientific productivity, the dependent variable, was defined as the number of scientific publications. A word should be said about the dependent variable. It is scarcely unassailable. A scientist's professional publications are one aspect of his scientific role, and one of his contributions to his professional group and to the greater society of which he is a part. Quality of publications, although difficult to assess, was not ignored, for publications in professional journals are reviewed by editorial boards which maintain that articles "must make a significant contribution to an area of psychology [Shaffer, 1957, p. 6]." To preserve quality considerations, articles were counted only if they appeared in journals maintained by established professional associations, with professional editorial consultants, and national circulations. The choice of this dependent variable, from among the many measures of therapeutic outcome, does not deny the importance of other criteria, but it may be recommended, perhaps, because of its relatively public nature. Although it is important to study alterations in self-perception as a result of therapy, it is equally important to investigate objective behavioral consequences.

Three hypotheses were tested. The first hypothesis is that the mean after-therapy productivity of a sample of psychologists who have received therapy is higher than the mean productivity of a matched nontherapy control sample. Psychotherapy may, however, produce differential effects depending upon the individual's productivity history. A relatively nonproductive individual may be freed from intrapsychic conflicts, better able to utilize his productive capacities, and may become more productive after therapy (Kris, 1952; Kubie, 1958). On the other hand, after therapy an individual who had been "driven" to high productivity may be less compulsive than Snow (1961, p. 72), for example, deemed proper for scientists, and he may produce less. If, after therapy, the more productive psychologists write less, while the less productive ones produce more, the mean productivity of the therapy sample will not be significantly different from the nontherapy control sample, but therapy still had an impact. The second hypothesis, therefore, is that the correlation between before-therapy and after-therapy productivity for the therapy sample will be negative and significantly different from the "earlier" and "later" productivity of the nontherapy control sample. Finally, successful psychotherapy may "unfreeze" existing productivity patterns so that both productive and nonproductive psychologists may become either more or less productive after therapy. In other words, successful therapy would make for greater freedom of movement. The third hypothesis therefore, is that the mean of the absolute differences between the amount of productivity before therapy and after therapy of a sample of psychologists who have received therapy is greater than the mean of the absolute differences between the earlier and later productivity of a matched nontherapy control sample.

---

[1] This study was completed while the senior author was at the Laboratory of Psychology, National Institute of Mental Health. The author is obliged to Joseph M. Bobbitt for his support and cooperation.

METHOD

A questionnaire was used to obtain a sample of psychologists who had received psychotherapy within the time intervals, and of sufficient duration, to fulfill the requirements of this study, and to provide information for matching a nontherapy control group. The questionnaire asked for the institutions and dates of the baccalaureate and doctoral degrees (Levy, 1962), the areas of specialization, the professional activities (research, teaching, counseling, administration, etc.) and the principal employers for the "first five years after receipt of Ph.D.," and for the "last five years." The respondents were also asked whether, during the past 15 years, they had had, or had considered, psychotherapy, and their reasons for, or against, entering it.

Those respondents who had received therapy were asked the dates of beginning and termination of therapy, and the orientation of the therapist (psychoanalyst, psychotherapist, clinical psychologist, social worker, clergyman). They were also asked to evaluate on a seven-point scale the effects of psychotherapy on "your ability to see yourself as others see you," "to understand your underlying motives," "your feelings of comfort," "professional effectiveness" (Parloff, Kelman, Frank, 1954), and the "effect of psychotherapy on the quantity and quality of your professional writing."

Questionnaires were sent to all APA members who had received their doctorates from American universities 1945–1951, inclusive, but in order to maintain as much sample homogeneity as possible, women psychologists were eliminated from the analyses. A total of 1,123 questionnaires were sent to male psychologists, of which 86% were returned. Contingency analyses of a 50% sample of nonrespondents showed that the no-bias hypothesis was tenable for: (*a*) age, (*b*) year of PhD, (*c*) geographic location of the doctoral university, and (*d*) APA divisional memberships, all information being obtained from the 1960 APA *Directory* (Holsopple, 1960). There was also no significant difference between respondents and nonrespondents in their mean yearly weighted productivity scores.

The mean yearly weighted productivity score was obtained for each therapy and control respondent by counting his publications listed in the *Psychological Abstracts*. Where the abstract entries were ambiguous, or where doubt existed for any other reason, the books or articles were checked directly. Following previous investigations (Meltzer, 1949; Wispé, 1965), an empirical weighting scheme was used. A book written alone was given a weight of 36 points, a book with one collaborator, 18, a book with two collaborators, 12, a revision of a previously published book, 18, a book edited, 2, a chapter in a book, 2, an article written alone, 2, an article written in collaboration, 1, various reviews, notes, etc., 1.

For each therapy respondent, before-therapy and after-therapy mean yearly weighted scores were obtained. The former included the weighted score for all articles published from 2 years prior to the respondent's PhD through his first year of therapy, divided by the number of years between his PhD and the beginning of therapy; while the latter included the mean weighted score for all publications from the first year after therapy through 1961. A comparable score was obtained for each of the matched controls in the manner described below.

For each therapy respondent, a control respondent was matched on the basis of: age or date of PhD (whichever was possible), type of professional activity, area of specialization, and earlier productivity, the score of the control being within the same quartile as the before-therapy productivity score of the therapy respondent with whom he was matched. In matching the therapy and the control respondents the number of years sampled was kept the same. For example, if the therapy respondent received his doctorate in 1950, entered therapy in 1954, and terminated in 1957, his before-therapy productivity score was obtained for the period 1948–1954 and his after-therapy score for the years 1958–1961. If his matched control received his doctorate in 1951, his earlier mean yearly weighted productivity score was obtained for 1949–1955, and his later score for the years 1959–1961. Contingency analyses for the variables above showed no significant differences between the therapy and the control samples.

In order to insure reliable measures of before-therapy and after-therapy productivity, it was necessary that therapy begin not less than 2 years after receipt of the PhD and terminate by December 1958. It should be noted that amount and type of professional experience, important factors in scientific productivity, were controlled by the matching procedures.

It was possible to obtain a sample of 55 male psychologists who had received 60 hours or more of psychotherapy, this being the arbitrary minimum used to define the therapy group, and for whom 55 matched psychologists could be found who had neither considered, nor received, therapy. It was with this group of 110 psychologists that the main hypotheses were tested.

RESULTS AND DISCUSSION

Since the respondents appeared representative of the 1945–1951 doctorates who were members of the American Psychological Association, we address first the question: What proportion of American male psychologists have had psychotherapy, and who are they?

Of the 966 psychologists who returned questionnaires, one-third had received psychotherapy at one time or another. Of the two-thirds who had not received any psychotherapy, 22% indicated they "had considered therapy sometime during the last 15 years," but had not entered therapy for reasons that ranged from, "my problem did not warrant therapy" to "lack of funds," "lack of motivation," and the substitution of "self-analysis," in that order.[2] These reasons are strikingly similar to those reported in a national survey of mental health practices where people who felt a need for psychological help but did not get it gave as their reasons, "lack of knowledge about where to go," "shame and stigma," and "the desire to work the problem out for themselves [Gurin, Veroff & Feld, 1960, p. 362]."

Of the 316 respondents who had received psychotherapy sometime during their lives, 172 respondents had received more than 60 hours of therapy, although for various reasons they could not be used to test the major hypotheses. These respondents were used to determine what, if any, characteristics obtainable by questionnaire distinguished psychologists who have had psychotherapy. When this sample of 172 psychologists who had, at any time in their lives, received more than 60 hours of therapy was compared with a representative random sample of nontherapy respondents who had obtained PhDs from American universities 1945–1951, there were no significant differences in the fathers' occupational levels, nor in the fathers' and mothers' educational levels. There were also no geographical significant differences in the universities from which the respondents received their doctorates, although more of the therapy sample received PhDs from private universities ($\chi^2 = 7.47$, df $= 1$, $p < .01$). Comparison of the areas of specialization showed more clinicians in the therapy sample and more industrial and educational psychologists in the random nontherapy sample ($\chi^2 = 41.55$, df $= 9$, $p < .01$). This overrepresentation of clinicians in the therapy sample and educational and industrial psychologists in the nontherapy sample became more pronounced between "the first five years after the Ph.D." and "the last five years," and the difference carried over into types of employment. More respondents in the therapy sample worked in clinics, private practice, and government nonresearch setting ($\chi^2 = 26.14$, df $= 11$, $p < .01$). There was no difference, however, between this therapy sample and the random nontherapy sample in their total research productivity.

When asked to indicate their "reasons for

TABLE 1

MEANS, MEDIANS, STANDARD AND QUARTILE DEVIATIONS OF THE BEFORE-THERAPY AND AFTER-THERAPY MEAN WEIGHTED PRODUCTIVITY SCORES

	M	*SD*	Mdn.	*Q*	Tau odd-even reliability coefficients
Before-therapy productivity of therapy group ($N = 55$)	1.40	1.67	.70	.85	$.37^{**}$
Initial productivity of control group ($N = 55$)	1.33	1.60	.75	.97	$.37^{**}$
After-therapy productivity of therapy group	1.41	2.14	.43	1.20	$.38^{**}$
Final productivity of control group	1.14	2.04	.40	0.55	$.46^{**}$

$^{**}p < .01$.

---

[2] Approximately these same percentages applied to the sample of responding women psychologists.

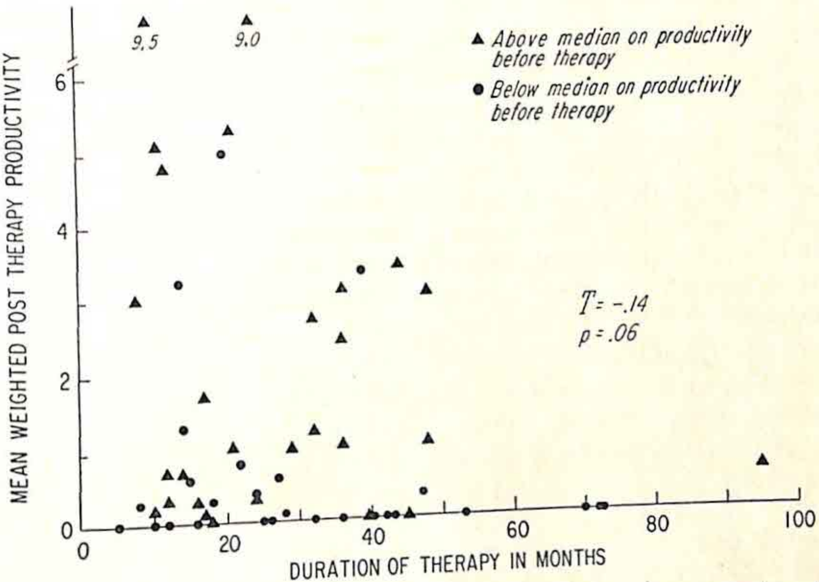


Fig. 1. Scatter diagram of the after-therapy productivity and the duration of therapy for respondents above and below the before-therapy median.

entering psychotherapy," 65% of this therapy group said they entered therapy for "personal" rather than "professional" reasons. When asked to indicate the theoretical orientations of their therapists, 69% of the 172 respondents checked psychoanalytically oriented therapists, 16% indicated psychiatrists with more eclectic orientations, and 12% said they had received therapy from a clinical psychologist, regardless of his theoretical bent. It is interesting to note that 41% of this group reported more than one therapeutic experience, and 13% had 3 or more.

Table 1 presents the means, medians, standard and quartile deviations, and the tau odd-even reliability coefficients for the therapy and the control samples on which the major hypotheses were tested. The reliability coefficients were all highly significant ($p < .01$), but at the lower limits for acceptability. Table 1 also shows that both variables were skewed toward the low end of the scale. Turning now to the major hypotheses, the impact of psychotherapy on productivity, Table 1 shows that although the mean and median after-therapy productivity of the therapy sample were slightly higher than the control sample, the differences were not significant. The tau's for the before- and after-therapy productivity for the therapy sample ($T = +.52$) and the earlier and later productivity for the control sample ($T = +.59$) were highly significant ($p < .01$), but not in the direction predicted by the second hypothesis. There was also no significant difference between the means of the absolute differences between the before- and after-therapy productivity of the therapy sample and the earlier and later productivity of the control sample. Psychotherapy did not appear to affect scientific productivity in a statistically significant manner.[3]

Figure 1 shows the scatter plot for duration of therapy and after-therapy productivity, with the before-therapy productivity indicated as above or below the median. The tau between after-therapy productivity and duration of therapy was low and not significant using a two-tailed test ($T = .14$, $p < .06$). Figure 1 shows, too, that with one exception the highest producers after therapy were above the median on productivity before therapy, and that most of the lowest producers after therapy were below the median on before-therapy productivity.

---

[3] Chi-square analyses failed to show any significant relationships between after-therapy productivity and: (a) reasons for entering therapy, (b) orientation of therapist, (c) field of professional specialization, and (d) the comparative education level of the parents, when this was dichotomized into those respondents whose mothers had the same or more education than the fathers, and those respondents whose fathers had more education than the mothers.

There was no relationship between before-therapy productivity and duration of therapy in months. Since productivity level before therapy was also unrelated to one's reasons for entering therapy, the data do not support the notion that respondents entered therapy, or received extensive therapy, because they were either unproductive or highly productive.

The items relating to the respondent's appraisal of his "objectivity," "insight," "comfort," and "competence" after therapy were highly intercorrelated, and were summed into a Therapy Satisfaction Index. The Therapy Satisfaction Index and the questionnaire items asking the respondents whether therapy affected their publications were correlated +.95. On the other hand, there was no significant relationship between either of these two items, or the Therapy Satisfaction Index, and the objective gains in quantity of publications as measured by the respondents' productivity scores! This interesting lack of relationship casts suspicion upon any unvalidated items which ask respondents to indicate improvements in the quantity of their work.

The relation between the Therapy Satisfaction Index and duration of therapy in months was positive and highly significant ($\chi^2 = 9.48$, $df = 4$, $p = .05$). The more therapy the respondent had, the more favorably he evaluated his therapeutic outcome.[4]

CONCLUSION

It appears that about one-third of the more recent doctorates in American psychology have at one time or another received some psychotherapy, and that a disproportionate number of them were, or later became, clinicians, working in a clinical setting. Although they sought therapy more for "personal" than for "professional" reasons, they were neither more nor less productive than a random sample of their colleagues. Moreover, psychotherapy did not significantly affect the subsequent level of productivity of psychologists when comparison was made with a carefully matched nontherapy control group.

The lack of statistically significant differences between the therapy and the control samples may be related to the increased variability of the therapy sample on after-therapy productivity. The middle 50% of the therapy sample covered a range on the after-therapy productivity scores about twice as great as the middle 50% of the control sample. This finding is the more interesting since the two samples were controlled on those sources of variance one would expect to be highly correlated with subsequent productivity. Two factors may explain this increased variability on after-therapy productivity. In the first place, an examination of Figure 1 shows that none of the five highest after-therapy producers had extensive therapy, and none of the five respondents with the most extensive therapy were highly productive afterwards. This suggestion of a negative relationship between duration of therapy and after-therapy productivity is strengthened by the fact that although the reported tau was low and not significant, none of the partial correlations (using tau) changed the sign of this relationship (Kendall, 1955). For example, holding constant the number of years since therapy terminated, the partial tau between duration of therapy and after-therapy productivity was still −.09. In the second place, the data show that more of the therapy sample than the nontherapy sample who were above the median on before-therapy productivity maintained, or improved, their subsequent productivity. Table 1 shows, too, that the therapy sample maintained its productivity level somewhat better than the nontherapy sample. Thus the therapy sample contained some respondents who wanted, or needed, and received, extensive therapy, and these respondents, although functioning, were not productive before therapy and were no more productive after therapy. The therapy sample, however, also contained some respondents who were very productive before therapy, and who were as productive, or more productive, after therapy. These two divergent subgroups account for some of the increased variability on the after-therapy productivity of the therapy sample as compared with the nontherapy sample. This analysis also implies that "having had therapy" is not a unitary variable.

---

[4] There were no significant relationships between the Therapy Satisfaction Index and: (*a*) reasons for entering therapy, (*b*) the orientation of the therapist, nor (*c*) the fields of professional specialization. The data failed to confirm the widespread notion that clinical psychologists, because of their professional orientation and bias, would evaluate their therapeutic experience more favorably.

All of this suggests that psychotherapy may have had some effect, even if it failed to reach statistical significance. The impact of therapy upon productivity may have been different for different individuals, depending, for example, upon the importance of scientific publications in the adjustment of the individual. Further investigation is needed to test the hypothesis related to productivity, not that therapy frees creativity or reduces compulsivity, but that it enables those who are already productive to maintain their productivity when, for various reasons, their colleagues' output is declining.

REFERENCES

GURIN, G., VEROFF, J., & FELD, SHEILA. *Americans view their mental health*. New York: Basic Books, 1960.

HOLSOPPLE, J. Q. (Ed.) *American Psychological Association 1960 Directory*. Washington, D. C.: APA, 1960.

KENDALL, M. G. *Rank correlation methods*. New York: Hafner, 1955.

KRIS, E. *Psychoanalytic explorations in art*. New York: Bailey, 1952.

KUBIE, L. S. *Neurotic distortion of the creative process*. Lawrence: Univer. Kansas Press, 1958.

LEVY, L. The skew in clinical psychology. *American Psychologist*, 1962, **17**, 244–249.

MELTZER, B. N. The productivity of social scientists. *American Journal of Sociology*, 1949, **55**, 25–29.

PARLOFF, M., KELMAN, H., & FRANK, J. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *American Journal of Psychiatry*, 1954, **8**, 343–351.

SHAFFER, L. F. (Ed.) *Publication manual of the American Psychological Association*. Washington, D. C.: APA, 1957.

SNOW, C. P. *Science and government*. Cambridge: Harvard Univer. Press, 1961.

WISPÉ, L. G. Some social and psychological characteristics of eminent psychologists. *Journal of the History of the Behavioral Sciences*, 1965, **1**, 88–98.

(Received February 24, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 194–200

A BEHAVIORAL STUDY OF STAFF ATTITUDES TOWARD MENTAL ILLNESS[1]

ROBERT B. ELLSWORTH[2]

Veterans Administration Hospital, Roseburg, Oregon

65 nurses and aides completed 2 factor analyzed attitude questionnaires. Hospitalized psychiatric patients (N = 188) screened for ability to identify their staff and make valid behavioral ratings, rated their ward staff on a 55-item interpersonal behavior inventory. Relationships between perceived behavior and endorsed attitudes were examined. Restrictive attitudes were consistently related to controlling and restricting behavior. Protective Benevolence was related to such behaviors as aloofness, distance, and dishonesty. Hence, some attitudes were related to behaviors which were congruent with the endorsed attitude, some attitudes were related to seemingly incongruent behaviors, while other attitudes had no significant behavioral correlates. A new attitude area, highly related to outgoing interpersonal behavior, was identified.

One of the basic assumptions underlying recent attempts to define and measure staff attitudes toward mental illness is that differences in attitudes do, indeed, make a difference in one's behavior and effectiveness in working with psychiatric patients. This assumption has received some support (Carstairs, Gilbert, Heron, Levinson, & Pine, 1957; Pine & Levinson, 1957). The attitude questionnaire items used in these earlier studies were scored along a single dimension of "custodialism-humanitarianism." Later investigators have suggested that opinions about mental illness are multidimensional, and have applied the technique of multifactor analysis to identify these attitude dimensions.

It was the purpose of the present study to determine whether or not one's endorsement of these newer attitude dimensions made any significant difference in one's behavior in relating to psychiatric patients. Sixty-five aides and nurses completed two attitude questionnaires and were scored on 16 attitude dimensions (5 Opinions about Mental Illness (OMI) factors, Cohen & Struening, 1962; 6 Staff Opinion Survey (SOS) factors, Berger, 1962; and 5 combination OMI-SOS factors, Struening, 1963). Psychiatric patients were carefully screened both for their ability to recognize and correctly identify the pictures of their ward personnel as well as their ability to handle a task requiring the rating of interpersonal characteristics. A total of 188 patients (out of 382 screened) rated the behavior of the ward personnel they knew, on a 55-item interpersonal rating scale.

The relationships between staff attitudes and staff behavior were computed. Only 7 of the 16 attitude dimensions were related to any significant numbers of staff behavior. Staff people who endorsed 1 of the 5 authoritarian-control-restriction attitude dimensions were seen by patients as behaving significantly more often in a controlling, restricting, and domineering way, as compared with staff members who rejected these attitude dimensions (i.e., OMI Factors A and D, SOS Factor I, and OMI-SOS Factors 1 and 3). The second major finding was that the staff member who endorsed Protective Benevolence (OMI-SOS Factor 4) was seen as behaving more frequently in an aloof, distant, and cold manner. And finally, it was discovered that many of the same behaviors (rated by patients) were related to *both* of these staff attitude domains. Both staff members who endorsed restriction and control, or benevolence, were rated more frequently as showing a lack of respect in their behavior toward patients (i.e., not honest or dependable, treats patients not as equals but more like children, etc.). From the behavioral standpoint, then, the 116 OMI-SOS attitude items seemed to tap at least two basically different, but somewhat overlapping, attitude dimensions.

---

[1] This study was supported by VA 8200 Research Funds.

[2] In cooperation with Leslie Douglass and Maxine Mattox, who were responsible for most of the testing, scoring, and statistical computations.

SHARPENING THE ATTITUDE DIMENSIONS

The two general attitude areas of restriction and benevolence apparently do make a difference in one's behavior toward patients. It was necessary, however, to sharpen these attitude dimensions in order to reduce the behavioral overlap between them. OMI-SOS Factor 3 (Restrictive Control) and Factor 4 (Protective Benevolence) were selected for further analysis since these two factor scores were differentially related to the two behavioral clusters characterized by control and restriction versus aloofness and distance. In order to sharpen and clarify these attitude dimensions which underly differences in staff behavior, the relationships between the scores of OMI-SOS Factors 3 and 4 and all 116 attitude items were computed. This item analysis was first completed on the attitude questionnaires of 106 aides, and cross-validated on an independent sample of 158 relatives. Fifteen attitude items were found to be significantly correlated with the OMI-SOS Restrictive Control scores for both the aide and relative populations. Ten attitude items were found to be significantly related to the OMI-SOS Protective Benevolence scores. An additional 14 attitude items were found to be significantly related to the scores of *both* Restrictive Control and Protective Benevolence. Thus, a third attitude dimension emerged. As will be seen later, it was this third attitude dimension which largely accounted for the overlap in behaviors of those staff members who endorsed either the original Restrictive Control or Protective Benevolence attitude dimensions. This third attitude area appears to tap an attitude dimension best labeled as "Nontraditionalism."

RESULTS AND DISCUSSION

Restrictive Control

As noted above, there were 15 attitude statements which were significantly correlated with the original OMI-SOS Factor 3 attitude scores for both the aide and relative population. These 15 attitude statements are:

1. Although patients discharged from mental hospitals may seem all right, they should not be allowed to marry.
2. All patients in mental hospitals should be prevented from having children by a painless operation.
3. People with mental illness should never be treated in the same hospital as people with physical illness.
4. A woman would be foolish to marry a man who has had a severe mental illness, even though he seems fully recovered.
5. It is easy to recognize someone who once had a serious mental illness.
6. If a patient in a mental hospital attacks someone, he should be punished so he doesn't do it again.
7. If a mental patient is allowed to keep his watch and jewelry, it creates more trouble than it is worth.
8. The law should allow a woman to divorce her husband as soon as he has been confined in a mental hospital with a severe mental illness.
9. Mental patients come to the nursing staff with too many unimportant problems.
10. The small children of patients in mental hospitals should not be allowed to visit them.
11. People would not become mentally ill if they avoided bad thoughts.
12. It is not a good idea for the mental patient to stay on one ward as long as he is in the hospital.
13. Most patients returning from a leave should be searched for forbidden items.
14. Anyone who is in a hospital for a mental illness should not be allowed to vote.
15. Mixing of men and women patients should be discouraged.

In general, this attitude dimension reflects an endorsement of restriction, control, segregation, and possible punishment in the handling of mental patients. There is also a feeling of pessimism in the outlook for recovery since mental illness is looked upon as a rather

TABLE 1

PATIENT PERCEPTION OF STAFF WHO ENDORSE THE ATTITUDE OF RESTRICTIVE CONTROL

Staff behavior (patient perception)	r with attitude score
Sensitive and understanding	−.50
Impatient with others' mistakes	+.47
Hard-boiled and critical	+.46
Open and honest with me	−.41
Treats me as an equal	−.40
Skeptical and suspicious	+.37
Trusts patients	−.35
Finds time to listen to me	−.35
Efficient and business-like	−.35
Laughs at a joke	−.34
Cold and unfeeling	+.34
Avoids patients	+.34
Goes out of his way to help me	−.33
Bossy and domineering	+.32

Note.—The 14 behavior items above are related to attitude scores at the .01 level of significance. An additional 13 behavior items were related to this attitude at the .05 level of significance.

permanent condition. These 15 attitude items are highly similar in content to the attitude dimension tapped by Cohen and Struening's (1962) Social Restrictiveness Factor D, although only one-third of the 15 attitude items in the modified Restrictive Control area are included in the OMI Factor D. The statistical relationship between Cohen and Struening's OMI Factor D and the new Restrictive Control area was found to be very high (i.e., .91).

As seen in Table 1, the staff member who endorses the attitude of Restrictive Control was seen by patients as a somewhat inconsiderate, impatient, rigid, domineering, skeptical person who was relatively unable to trust, relate to, or understand patients. He was also perceived as not "efficient and business-like," which was discovered in later interviews with patients to describe a person who does not take care of requests and problems. Only those 14 behaviors whose correlations[3] with the attitude score were significant beyond the .01 level are presented in Table 1.

The present finding of the relationship between an endorsement of the attitude of Restrictive Control and the frequency of perceived restrictive staff behavior has already been noted in the work of Carstairs et al. (1957). In their study, it was found that the aide who endorsed Custodial attitudes was rated by his supervisor as handling interpersonal relationship problems with patients by control, restriction, and punishment. In another study, Cohen and Struening (1964) found that hospitals whose personnel endorsed attitude combinations of authoritarianism-restrictiveness were not particularly effective in returning patients to the community. Apparently, then, the attitude of restrictiveness does make a difference in one's behavior, and may also make a real difference in how effectively one works with hospitalized patients in returning them to the community.

Protective Benevolence

There were 10 attitude statements which were significantly correlated with the original OMI-SOS Factor 4 attitude scores for both the aide and relative populations. These 10 attitude statements are:

1. Patients should never be locked up alone in an isolation room.
2. Our mental hospitals should be organized in a way that makes the patient feel as much as possible like he is living at home.
3. People (both veterans and nonveterans), who are unable to work because of mental illness, should receive money for living expenses.
4. The staff should be as friendly with patients as they are with one another.
5. Patients should be paid for any work they do in the hospital.
6. A patient should be placed on a ward where most other patients are like him in age, education, and type of illness.
7. If a mental patient does not like his work assignment, he should be allowed to change it.
8. Mental patients who cause the least trouble in the hospital are likely to get along well after discharge.
9. Most mental patients are willing to work.
10. Nursing attendants should try to make friends with their patients.

In general, the Protectively Benevolent person believes that a patient should be made to feel like he is living at home, that he should be paid for work he does and should receive money for living expenses, that he should be allowed to change his work assignments, and that he should be approached in a friendly, nonpunitive way.

Table 2 presents the behavior which patients perceive as portraying the staff person who endorses the 10 item attitude area of Protective Benevolence. It is apparent from this table that the Protective Benevolent staff person behaves in ways which are inconsistent with the attitude he endorses. As seen in Table 2, the Protectively Benevolent staff

---

[3] The 188 patients who rated staff behavior resided in a 6 unit, 3 floor, psychiatric building. The number of nursing personnel rated by each unit's patients was 11, 11, 9, 12, 12, and 10 (total $N = 65$). For each unit, rank order correlations between staff attitude scores and each of the 55-patient rated behaviors were computed, converted to Fisher z scores, combined, and averaged. Only correlations significant beyond the .01 level (i.e., .32 or better for 63 *df*) are reported in order to reduce the possibility of reporting relationships which could have arisen by chance. The approach of analyzing each unit separately, and combining the resultant correlations was the approach which seemed most desirable since each unit was different; i.e., emphasized different degrees of patient control, permissiveness, treatment philosophy, etc. Empirically, however, this approach to data analysis yielded essentially the same results as compared with overall product-moment correlations based on the total N of 65.

TABLE 2

PATIENT PERCEPTION OF STAFF WHO ENDORSE THE ATTITUDE OF PROTECTIVE BENEVOLENCE

Staff behavior (patient perception)	r with attitude score
Stays by himself	+.45
Lets patients get to know him	−.39
Talks about a variety of things	−.38
Honest, straightforward with me	−.36
Keeps his promises to me	−.36
Patients turn to him for advice and help	−.35
Talks a lot	−.34
Gives freely of himself	−.33
Reserved and cool	+.33
Goes out of his way to help me	−.33
Accepts suggestions from me	−.32

Note.—The 11 behavior items above are related to attitude scores at the .01 level of significance. An additional 8 behavior items were related to this attitude at the .05 level of significance.

member is perceived by patients as a basically aloof, distant, noninteracting kind of person. One must, therefore, interpret the attitude items of Protective Benevolence as communicating something other than "kindliness toward patients" in order to reconcile this inconsistency. What the Protectively Benevolent person may really be saying is that he finds it preferable to indulge, placate, and generally go along with and accept the patient as he is. For example, in talking with staff people who scored high in Protective Benevolence, there emerges a philosophy of patient handling which is best summarized as: "It's better to leave the patient alone because if you confront him he just gets upset." Apparently many Protectively Benevolent staff people tell the patient what he wants to hear, which in turn is perceived by patients as (not) "honest and straightforward with me," and (not) "keeps his promises to me" (see Table 2). Staff people, for example, may tell a demanding patient he will get a privilege card, not because they necessarily believe he deserves one but because it may be useful in temporarily reassuring a disturbed patient. Other examples of this placating, but perceived dishonest behavior, can readily be found in the actions of the Protectively Benevolent mother or wife who writes the staff that she does not want the patient home, but writes the patient that she does want him home. Upon examination, she usually reveals that she does not want to "hurt the patient's feelings," and chooses to engage in this inconsistency at the expense of honesty.

A basic goal of many Protectively Benevolent staff people, as suggested by the examples above, is probably one of maintaining a superficially comfortable relationship with the patient by accepting the patient as he is and avoiding confrontation. What the Protectively Benevolent staff person may well be describing is exactly this kind of relationship when he endorses such ideas as "make the patient feel like he's living at home," "pay the patient for work he does," "allow him to change his work assignment," "make friends with patients," etc. If, on the other hand, one's primary motivation is to help the patient to change, then change often demands confrontation. Confrontation carries with it the risk of stress, differences of opinion, and discomfort for both the patient and the staff. In short, it may be easier for the Protectively Benevolent staff person to remain aloof and distant from the patient in order to maintain, as his first priority, a comfortable relationship.

Protective Benevolence, as measured in this study, is *not* the same attitude dimension as defined by the OMI Benevolence factor (Cohen & Struening, 1962). First of all, the OMI Factor B was not consistently related to any perceived differences in staff behavior. Secondly, the correlation between the OMI Benevolence and the present dimension of Protective Benevolence was only .34. As Cohen and Struening define their dimension, it would seem that their factor would have been described better as Humanistic Benevolence since it seems to reflect in part a feeling of "Christian kindliness toward unfortunates." The present study did not find that even this humane attitude (OMI Factor B) was consistently related to positive, warm, friendly staff behavior. A lack of clear-cut findings would lead one to suspect that Cohen and Struening's Benevolence items were probably endorsed by both genuinely humanistic staff people, and as found in the present study, by staff who basically wish to establish a comfortable but aloof relationship, and who rationalize their behavior as an act of kindness.

Nontraditionalism

During the item analysis of the aide and relative questionnaires, 11 attitude items were found to correlate significantly with the original scores of *both* Restrictive Control and Protective Benevolence. In other words, there was a common core of attitude statements which both Restrictive Control and the Protectively Benevolent staff endorsed and accepted. An additional 3 items were later added through an item analysis in which the relationship of all 116 attitude items and the new attitude area was determined. These 14 attitude items, rejected by the nontraditional staff member, are:

1. One of the main causes of mental illness is a lack of moral strength or will power.
2. When a person has a problem or worry it's best not to think about it, but keep busy with more pleasant things.
3. Although some mental patients seem all right, it's dangerous to forget for a moment that they are mentally ill.
4. People who are successful in their work seldom become mentally ill.
5. More mental patients should be committed by the courts so that the hospital is really in charge.
6. Mental illness is usually caused by some disease of the nervous system.
7. Mental patients come from homes where the parents took too little interest in their children.
8. Sometimes mental illness is a punishment for bad deeds.
9. If parents loved their children more, there would be less mental illness.
10. Most patients in mental hospitals don't care how they look.
11. There is something about mental patients that makes it easy to tell them from normal people.
12. Every mental hospital should be surrounded by a high fence and guards.
13. People who are mentally ill let their emotions control them; normal people think things out.
14. Nervous breakdowns usually result when people work too hard.

The dimension tapped by these attitude statements seems relatively clear. In general, the nontraditional staff member rejects such traditional concepts as the cause of mental illness is found in lack of will power, disease of the nervous system, and too little parental interest or parental love. The nontraditional staff member also rejects the idea that patients are unpredictable and dangerous, and that patients are different in appearance, motivation, and emotions than a normal person.

As seen in Table 3, there are 35 behavioral characteristics which the patient perceives in his relationship with the nontraditional staff member. This attitude dimension was by far the most potent and significant dimension arising out of the present study. Not only was it significantly related to more behaviors than any other attitude dimension, but the size of the correlations was consistently greater.

The explanation of why this attitude dimension was so highly related to staff behavior is not entirely clear. Perhaps it is a measure of whether or not staff people believe that a patient is capable of change. If one believes that the origin of a patient's illness and behavior lies in the patient's past, the presum-

TABLE 3

PATIENT PERCEPTION OF STAFF WHO ENDORSE THE ATTITUDE ON NONTRADITIONALISM

Staff behavior (patient perception)	*r* with attitude score
Sensitive and understanding	+.68
Dependable and reliable	+.67
Open and honest with me	+.65
Gives advice freely	+.64
Honest, straightforward with me	+.60
Hard-boiled and critical	−.60
Bossy and domineering	−.60
Friendly and approachable	+.58
Impatient with others' mistakes	−.55
Says one thing but does another	−.55
Lets patients get to know him	+.52
Impatient and annoyed	−.52
Treats me like a child	−.52
Makes me feel at ease when talking to me	+.51
Teases or laughs at patients	−.51
Acts important	−.50
Goes out of his way to help me	+.50
Cold and unfeeling	−.47
Finds time to listen to me	+.46
Treats me as an equal	+.46
Keeps his promises to me	+.45
Skeptical and suspicious	−.44
Patients turn for advice and help	+.43
Keeps me informed	+.42
Shows a good sense of humor	+.40
Respects my opinions and ideas	+.40
Avoids patients	−.40
Talks about wide variety of things	+.37
Gives freely of himself	+.36
Trusts patients	+.36
Efficient and business-like	+.36
Looks out for my welfare	+.34
Is able to change his mind	+.34
Agrees with me	+.34
Laughs at a joke	+.33

ably one would conclude that a person, who cannot alter the origins of his illness and behavior, is unlikely also to be able to alter his present behavior. If, on the other hand, one rejects the concept of historical etiology, in which the patient is seen as a victim of circumstances beyond his control, then one is more apt to believe that patients can change, that they are treatable, and that interaction pays off. In addition, one may view the patient as dangerous, unpredictable, and different from normal people only if one feels that the patient is not directly responsible for his illness and subsequently not accountable for his behavior.

Another explanation of why this attitude dimension was so highly related to staff behavior is the possibility that it was simply a direct reflection of authoritarian values. The attitude area of nontraditionalism did correlate −.78 with the OMI Factor A (Authoritarianism). But it also correlated −.53 with Interpersonal Etiology (Factor E) and −.55 with Social Restrictiveness (Factor D). Since it was somewhat more highly related to the Authoritarian Factor A than any other factor, it would seem that the nontraditional attitude dimension does, in part, reflect authoritarian values.

As seen in Table 3, however, the main behavioral characteristics of the Nontraditional staff person are that he was seen as: "sensitive and understanding," "dependable and reliable," "open and honest," "gives advice freely," "honest and straightforward," etc. These elements appear to be more a measure of warmth and engagement with patients. Next in order, following the behavioral characteristics of warmth and engagement, are a rejection of those behavioral characteristics which are usually associated with authoritarian behavior ("hard-boiled and critical," "bossy and domineering," "impatient," etc.). In short, the staff person who rejects traditional concepts of mental illness is first of all able to engage in a warm and open way with patients. Perhaps the simplest explanation, then, is that the Nontraditional staff person believes that the patient is not the passive victim of circumstances beyond his control, that he is able to change, and that interaction pays off.

The 14-item Nontraditional attitude area is correlated − .46 with the 15-item Restrictive Control area and −.28 with the 10 item Protective Benevolence area. Restrictive Control and Protective Benevolence were found to be relatively independent of, and negatively related to, each other ($r - .21$). Thus, if a person endorses the concept of Nontraditionalism, he tends to endorse *neither* Restrictive Control nor Protective Benevolence. If one is more traditional in his attitudes toward mental illness, then one tends to endorse either Restrictive Control and/or Protective Benevolence. In behavioral terms, if one does not interact openly and warmly with patients, he can choose to control and restrict the patient, or to withdraw from the patient, as a rationalized act of benevolence.

As found in the present study, the three attitudes areas of Restrictive Control, Protective Benevolence, and Nontraditionalism are three basic attitude dimensions which make a difference in the behavior of psychiatric aides and nurses in relating to their patients. This study also suggests that attitude dimensions which are uncovered through factor analysis may represent mathematically cohesive ways of viewing mental illness, but that these factor dimensions do not necessarily make a difference in one's behavior.

The results of the present study lend support to the defining of an attitude as an underlying disposition which enters into the determination of a *variety* of behaviors toward an object or class of objects, including statements of beliefs and feelings about the object as well as one's actions with respect to it (Cook & Selltiz, 1964). A measure of one's verbal responses to an attitude questionnaire does not represent a complete measure of one's underlying attitude. Studies which report that student nurses, for example, change significantly in their attitudes as a result of a psychiatric affiliation (Gelfand & Ullman, 1961; Hicks & Spaner, 1962), are basically reporting a change in the manner in which the student nurses responded to a set of attitude statements toward mental illness. One cannot assume that a basic or underlying attitude change has indeed taken place unless one also knows the extent to which there has been a parallel change in the relationship behavior between the student nurse and the hospitalized patient.

And finally, the relationship between an employee's verbal attitude toward mental illness and his rated effectiveness in a hospital setting is dependent upon several factors. In a study by Toomey, Reznikoff, Brady, and Shumann (1961) no relationship was found between verbally expressed attitudes and success in psychiatric affiliation. Canter (1963), however, finds that performances and attitudes are related, but the extent of the relationship is dependent on what kind of attitude dimensions are being measured. In our experience, we have found that the extent to which attitudes and job performance are related also depends on the situational demands of a particular treatment ward. On a ward where the program demands that patients and staff relate to each other on a peer level, the staff person who is rated as most effective is one who endorses the Nontraditional attitude dimension. In other ward settings, however, this attitude is not an important one in differentiating the effective and the ineffective staff member. What is needed at this point is a serious attempt to evaluate the impact of different approaches on the treatment outcome of patients, and then to judge which attitudes are most critical in determining how well one functions in working with the more effective treatment approaches. It is premature to conclude that the employee who endorses Restrictive Control, for example, is less effective in helping patients than one who endorses Nontraditionalism. The final conclusions on the relationship between endorsed attitudes and effectiveness in patient rehabilitation will undoubtedly depend on what attitudes are being measured, the kind of demands of the treatment situation itself, and the kind of patient being treated.

REFERENCES

BERGER, DAVID & Others. Staff opinion survey. An unpublished factor analyzed attitude scale arising out of the Medical Audit Research, Veterans Administration Hospital, Perry Point, Maryland, 1962.

CANTER, F. M. The relationship between authoritarian attitudes, attitudes toward mental patients, and effectiveness of clinical work with mental patients. *Journal of Clinical Psychology,* 1963, **19,** 124–127.

CARSTAIRS, G. M., GILBERT, DORIS C., HERON, A., LEVINSON, D. J., & PINE, F. Ideology, personality, and role definition: Studies of hospital personnel. In M. Greenblatt, D. J. Levinson, & R. H. Williams (Eds.), *The patient and the mental hospital.* Glencoe, Ill.: Free Press, 1957. Pp. 197–208.

COHEN, J., & STRUENING, E. L. Opinions about mental illness in the personnel of two large mental hospitals. *Journal of Abnormal and Social Psychology,* 1962, **64,** 349–360.

COHEN, J., & STRUENING, E. L. Opinions about mental illness: Hospital social atmosphere and patient time in hospital. *Journal of Consulting Psychology,* 1964, **28,** 291–298.

COOK, S. W., & SELLTIZ, CLAIRE. A multiple-indicator approach to attitude measurement. *Psychological Bulletin,* 1964, **62,** 36–55.

GELFAND, S., & ULLMAN, L. P. Attitude changes associated with psychiatric affiliation. *Nursing Research,* 1961, **10,** 200–204.

HICKS, J. M., & SPANER, F. E. Attitude change and mental hospital experience. *Journal of Abnormal and Social Psychology,* 1962, **65,** 112–120.

PINE, F., & LEVINSON, D. J. Two patterns of ideology, role conception, and personality among mental hospital aides. In M. Greenblatt, D. J. Levinson, & R. H. Williams (Eds.), *The patient and the mental hospital.* Glencoe, Ill.: Free Press, 1957. Pp. 208–217.

STRUENING, E. L. An unpublished factor analysis of the combined OMI and SOS. Veterans Administration Hospital, Montrose, New York, 1963.

TOOMEY, LAURA C., REZNIKOFF, M., BRADY, J. P., & SCHUMANN, W. D. Some relationships between the attitudes of nursing students toward psychiatry and success in psychiatric affiliation. *Nursing Research,* 1961, **10,** 165–169.

(Received March 10, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 201–209

THE RING OF PUNISHMENT:

A THEORETICAL AND EXPERIMENTAL ANALOGUE OF REPRESSION-SUPPRESSION [1]

LEONARD WORELL [2]

Oklahoma State University

This paper considers how some punished behaviors become less accessible than other equally punished behaviors. A theory viewing repression and suppression as lying on a continuum with respect to the accessibility of punished behavior is developed around the construct of a "ring of punishment." This ring consists of continuing and related punishments for a series of behaviors; punished behaviors preceded and followed ("ringed") by other punished behaviors are comparatively inaccessible (repressed) while punished behaviors not so ringed are less inaccessible (suppressed). Experimental data supported the theory for associative learning that fell within the ring of punishment, but some discrepant results appeared for learning falling just outside the ring.

In 1893 Breuer and Freud (1924) first turned their attention to the concept of repression. In 1943 Sears discouragingly concluded that the experimental study of repression provided "no refinement to the (repression) theory, no addition of relevant new variables, no streamlined techniques that promise eventual solution of the problems posed by Freud [p. 120]." MacKinnon and Dukes' (1962) recent review of the literature on repression suggests that these statements are still applicable today. It has frequently been asserted that experimental work on repression suffers from artificially created experiences that are mild and transient, whereas clinically observed "repression" occurs in really meaningful and intense areas such as sex, hostility, etc. While there may be some merit to these considerations, the present view is that a major problem may lie in the conceptualization of the repression notion. In the following, a theory is proposed dealing with the antecedents to repression which simultaneously incorporates the concept of suppression.

Historically, Freud included both suppression and repression in his early work, but as his writings progressed the term suppression was gradually dropped. Other psychoanalysts, however, continued to use both terms and clove a distinction between the two which generally centered around "awareness." Thus, Alexander (1932) assigns an "unconscious inhibiting process" to repression, while suppression refers to the same process when it is carried out with conscious awareness. Apart from the analysts, Dollard and Miller (1950) also embrace a notion of awareness (here explicitly in terms of verbal behavior). They emphasize that repression is like suppression "except that it is much more strongly motivated and is automatic [p. 200]." What is meant by automatic is never made operationally clear. In any event, it is to be noted that none of these descriptive approaches has proven to be of heuristic value. This leads us to suspect that the fault lies in the inadequate statement of the problem.

A reading of psychoanalytic sources (Freud, 1925a, 1925b) reveals that there have been two major ways in which repression has been used. The first is what has been termed "primal" repression which refers to the "*exclusion* from consciousness" of material that is either painful or potentially painful. According to the present terminology, this form of repression refers to conditions where an individual has great difficulty in *acquiring* responses to particular (here punishing) stimuli. The second use of repression has been most commonly designated as "repression proper" which refers to the "*expulsion* from

---

[1] This work was supported by USPHS Grant M-4891 from the National Institute of Mental Health.

[2] The author thanks John Hirschlein for his assistance in data collection.

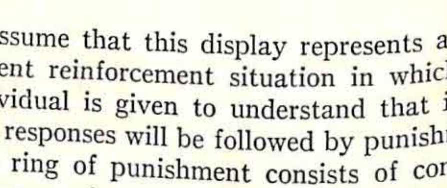
consciousness" of *already* acquired material because it has come to be associated with painful circumstances. The present paper is principally concerned with the first of these approaches.

In this account repression and suppression are continuous. Acceptance of this continuum means that we are dealing with a single dimension, so that one could discard either term and simply speak of the *degree* of repression or the degree of suppression. Since either designation is acceptable, the more important problem is that of defining operationally the nature of this single dimension. It is conceived here that the continuum is one of the *accessibility of punished behavior,* and we may refer to the degree to which punished behavior is accessible. With this view there is no need to suppose that verbal behavior, awareness or unconscious processes deserve a special place in relation to repression-suppression. Rather, all behavior, verbal or otherwise, may be seen in terms of accessibility, i.e., the likelihood of its occurrence with respect to particular stimulus events. Thus, we are saying that when one moves toward the repression end of the continuum we are dealing with comparatively inaccessible punished behavior, while moving in the opposite direction we view suppression as referring to relatively accessible punished behavior. Moreover, there may be a whole array of punished behaviors between these extremes which will have intermediate accessibility.

In view of the foregoing, the elemental issue that must be answered by any study purporting to deal with repression-suppression is—*how does it come about that behavior associated with some punishing or painful circumstances is less accessible than behavior associated with other (and often equivalent or stronger in intensity) punishing conditions?* As an example, we may take the hypothetical case of two children who have been punished with equal severity for lying behavior in the past. Following a particular episode in which an apparent lie has occurred, the parent confronts the child with the lie. One of the children may readily admit the behavior, while the other child may deny that it was a lie or even more deny that the behavior had ever taken place. What are the conditions producing the differential accessibility of these punished behaviors?

We propose to show that differential accessibility can be produced among a set of equally punished behaviors depending upon the position of each behavior within a *ring of punishment*. This "ring" basically consists of a series of *continuing* and *related* punishments. By continuing we mean that the potential or real punishment is always present in relation to particular stimuli, as for a child in the presence of an invariably punitive parent or for an infantryman in combat. Punishments may be related in two general ways: (*a*) by the continuous recurrence of the same punishment, as electric shock or a parent slapping a child and (*b*) on the basis of conventional principles of contiguity (hence temporally adjacent punishments) and generalization (where punishment occurs in the same or similar situations or where the deliverer of punishment is the same).

The following paradigm illustrates the ring of punishment, in which stimulus situation, response, and punishment are respectively represented by the letters S, R, and P:


$S_1 \rightarrow R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_4 \rightarrow$ terminates P
P—P—P—P—P—P
Ring of Punishment

Assume that this display represents a contingent reinforcement situation in which the individual is given to understand that incorrect responses will be followed by punishment. The ring of punishment consists of continually recurring electric shocks at 3-second intervals. With the occurrence of R_1 continuous intermittent punishment is instigated. Then R_2 is elicited but it does not terminate the punishment sequence so that R_3 is evoked. But this response is also ineffective so that finally R_4 is made which, although *followed by punishment itself* also provides for the escape from the ring of punishment because no further punishment follows.

In order to predict the accessibility of behaviors within the punishment ring, the following assumptions are made:

1. In punishment situations, the degree of accessibility of behavior, is contingent on the amount of avoidance, i.e., direct and generalized punishment effects. This assumption emphasizes the single dimensionality of repression-suppression.

2. Behavior that is unsuccessful in removing punishment is less accessible (more aversive) than behavior, although punished, that leads to escape from punishment.

3. In recurring punishment situations where the punishments are of equal intensity, the amount of avoidance is dependent on generalization gradients of avoidance and approach. The negative gradients are assumed to accrue from similar or adjacent punished responses. Approach gradients derive from similar or adjacent nonpunished responses and/or punished escape responses.[3]

To illustrate, let us take an individual who, in addition to experiencing the ring of punishment, has also used one behavior immediately before, and one behavior immediately after the ring which has been successful. Of the four responses within the punishment ring, the response with the least amount of avoidance is R_4. This lowered avoidance is attributable to two factors: (*a*) R_4 led to escape from punishment (Assumption 2) and (*b*) the generalization of approach from the response that was rewarded following the punishment ring (Assumption 3). On the other hand, the strongest quantitative avoidance will be associated with R_2 as a function of the fact that this response itself was unsuccessful and is further preceded and followed by unsuccessful responses R_1 and R_3. Hence, there will be a generalization of avoidance from both R_1 and R_3 that will increase the net strength of avoidance of R_2 above that of all other responses. By applying the same assumptions to the remaining responses, the *relative* strength of avoidance may be determined for each. The aversiveness of R_1 is reduced by the generalization of approach from the preceding nonpunished response. The aversiveness of R_3 is similarly reduced, but here by the generalization from the escape properties of R_4. The relative theoretical strengths for all responses are shown in Figure 1.[4]

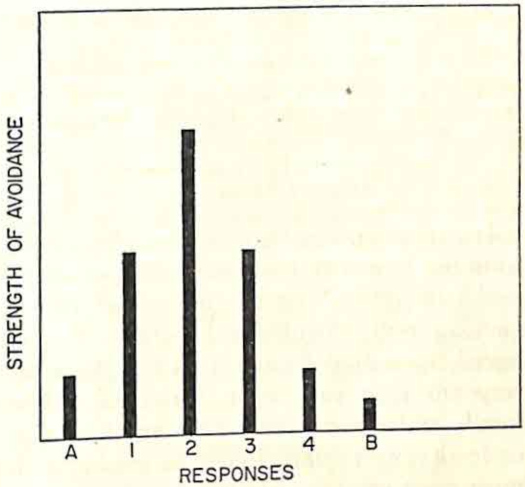


Fig. 1. Theoretical strengths of avoidance for responses within the ring of punishment and adjacent responses. (Numbered responses refer to punished responses; lettered responses signify rewarded ones.)

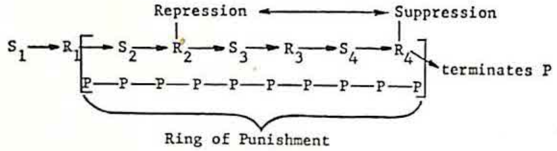
Repression-suppression may now be defined. A punished behavior is likely to be relatively inaccessible, i.e., "repressed," when it is both preceded and followed by punished responses that are unsuccessfully resolved, or ringed by punishment. In contrast, a punished behavior is most likely to be comparatively accessible, i.e., "suppressed," when it brings about the termination of a punishment sequence. It is important to note that toward the "suppressed" end of the continuum responses may or may not be inhibited in any given situation. If the situation elicits competing responses that are stronger, then the suppressed response will be inhibited. However, when the situation requires that the suppressed response be evoked, it will appear despite the fact that it has previously been associated with punishment.

An experiment is reported here which will use a subvariety of the general paradigm. This subparadigm consists of conditions in which the individual is responding to a multiplicity of stimuli within a general ring of punishment. It may be depicted as follows:

---

[3] Portions of this assumption find support in the work on spread of effect and temporal generalization (Church, 1963; Postman, 1962).

[4] The theoretical values in Figure 1 represent relative, and not absolute, strengths of avoidance. More precise values are contingent on either empirical knowledge of or the assumptions that one wishes to make about the slopes of the gradients of avoidance and approach.



It may be noted that this paradigm differs from the former in that many different stimuli evoke responses that are associated with an ongoing punishment. This paradigm is believed to be analogous to that of a child who responds to a variety of situations within a family setting but never does anything right, or to that of a bomber crew in combat which must react continuously to a bombardment of external stimuli with responses that do not reduce the threat. Grinker and Spiegel (1945) have offered a forceful picture of the variegated stresses (potential and real punishments) of a combat bomber crew, where such factors as the presence of flak or enemy aircraft, the loss of an engine, the wounding or killing of a buddy, possible defects in plane construction, the necessity to kill others (even civilians), the potential irresponsibility of other crew members, etc., constitute a genuinely harrowing ring of punishment. In the present study a parallel ring of punishment is created by providing continuing electric shock at 3-second intervals during a series of stimuli to which an individual must provide appropriate responses. According to the theoretical model developed, the repression end of the continuum will be represented by R_2, which has a greater inaccessibility relative to the remaining responses within the punishment ring. Since R_4, at the opposite end, despite being punished itself, is associated with successful escape behavior, it should be as accessible *when the situation requires* it as comparable behavior made to nonpunished stimuli.

METHOD

General Design

The experiment took place on two occasions. In the first part, each subject's upper level of electroshock tolerance was determined. Then, 48 highly meaningful nonsense syllables were presented to which free associations were made. At two predetermined points punishment was introduced and was continued at 3-second intervals through four successive stimulus-association pairs. When the fourth association was given, one additional shock was administered and the punishment sequence was terminated. The second part occurred 5 days later. In an initial free recall period subjects were asked to reproduce orally their previous associations and the nonsense syllables. Following recall, the nonsense syllables were presented in blocks of 12 syllables until the subject correctly anticipated the association that he had given to each syllable.

Equipment

Apart from a tape recorder, the major piece of apparatus was an electroshock dispenser, calibrated in 0.1-milliampere intervals, capable of delivering from 0.1 to 4.5 milliamperes of current. A cam attached to a 5-rpm motor activated a fulcrum that depressed the on-off switch at 3-second intervals. A counter was connected to the cam for recording the number of shocks administered.

Subjects and Procedure

A total of 36 volunteer males from introductory psychology sections was used. Each subject, participating individually, was asked whether he was willing to be in an experiment in which electric shock was employed. No subject refused. A commitment was then extracted from the subject that he would return 5 days later for the second part of the experiment. Information was also obtained from each subject about possible heart or other conditions that would interfere with his being in a shock experiment. Following these preliminary procedures, an assessment was made of each subject's pain-tolerance limit. Electrodes were attached to the subject's arm slightly above the wrist, and he was told that shock intensity would be gradually increased until he indicated that he could not tolerate any further increase. Up to an intensity of 1.0 milliampere, increases were made in steps of 0.1 milliampere, while beyond 1.0 the increases were in steps of 0.5 milliampere. Each subject's tolerance limit of intensity was used in the subsequent punishment sequences.

With the electrodes attached, the subject was told that nonsense syllables would be presented over a tape recorder and that he was to call out the first word that each one made him think of. He was further informed that he could not use the nonsense syllable as part of his association as, for example, TEX—Texas, Texaco, etc., and that he could not give the same association twice. The 48 nonsense syllables used (Table 1) were taken from the lists of Glaze (1928) and Krueger (1934) such that each syllable had a meaningfulness rating of 90% or above on both lists. Moreover, when the 48 syllables are separated into four groups of 12 (Table 1) the corresponding numbered syllable in each list has at least the first letter in common.

An attempt was made to motivate subjects by telling them that this word association test has been used as part of a battery of tests to select indi-

TABLE 1

Lists of Nonsense Syllables

List 1	List 2	List 3	List 4
TOS	TES	TAC	TIC
FIV	FEL	FOL	FUS
MED	MAK	MON	MOD
DUS	DIS	DEP	DEK
JIN	JON	JOL	JEL
SEM	SYN	SAR	SUT
BUZ	BAL	BEV	BEC
VIN	VIC	VAS	VOL
LUN	LUK	LOB	LEV
REV	RAV	RUS	ROV
COF	CUF	CEN	CUM
PIL	PIC	PUF	PUD

viduals for secret service assignments, and that studies at various universities had shown the association responses to be good indicators of basic scholastic achievement and the self-concept.

In order to establish the ring of punishment, subjects were instructed that some responses are considered fair, while others are good or even exceptional. Then,

> If you give a response that is rated as fair, you will receive a shock. The shock will continue while the next nonsense syllable is given and if your response to it is good, the shock will stop. However, if your response is fair, the shock will continue through the next nonsense syllable and response and so on until your responses are again rated as good. Just as soon as your responses are rated as good, the shock will stop until you give another fair response.

The nonsense syllables were presented sequentially through a tape recorder. Each syllable was first spelled and then pronounced, and the tape recorder was stopped until the subject gave a usable association.

In order to control for position and content effects of shocked and nonshocked nonsense syllables, four orders of punishment sequences were used with 9 subjects per order. These were:

Order	Punished sequences	
1	List 1—3, 4, 5, and 6	List 3—7, 8, 9, and 10
2	List 1—7, 8, 9, and 10	List 3—3, 4, 5, and 6
3	List 2—3, 4, 5, and 6	List 4—7, 8, 9, and 10
4	List 2—7, 8, 9, and 10	List 4—3, 4, 5, and 6

It may be noted that each subject received two series of intermittently occurring shocks, e.g., if the subject was in Order 1, he was shocked for the above numbered syllables and their associations in Lists 1 and 3. During the punishment sequences, shocks were delivered at 3-second intervals. In addition, the last association made in each punishment sequence, i.e., 6 and 10, was followed by one shock and then the shock sequence was terminated.

Five days later electrodes were again attached in order to reinstate the original situation. The electrodes were kept on throughout Part 2 but no shock was ever given. In a free-recall period the subject was asked to reproduce orally his previous associations and the nonsense syllables in two separate 5-minute intervals. Then, each subject was required to learn the associations that he had given to each syllable. The syllables were presented in four separate blocks of 12 syllables each through a tape recorder at 5-second intervals. Three seconds after each syllable was pronounced a tone was sounded. If the subject was unable to anticipate the association before the tone, the experimenter stated the association; if the subject anticipated correctly, the experimenter said nothing. Each list was repeated until the subject gave one complete set of 12 correct anticipations.

Results

Although the shock-tolerance limit accepted by subjects extended over a broad range from 0.8 to 4.5 milliamperes, all except three subjects took an intensity of 1.5 milliamperes or higher. The median tolerance limit value was 2.5 milliamperes.

Free Recall

Each subject was used as his own control in assessing the number of previous associations and nonsense syllables recalled. For example, if a subject received Order 1 where he was continuously shocked for his associations to Syllables 3, 4, 5, and 6 in List 1 and to

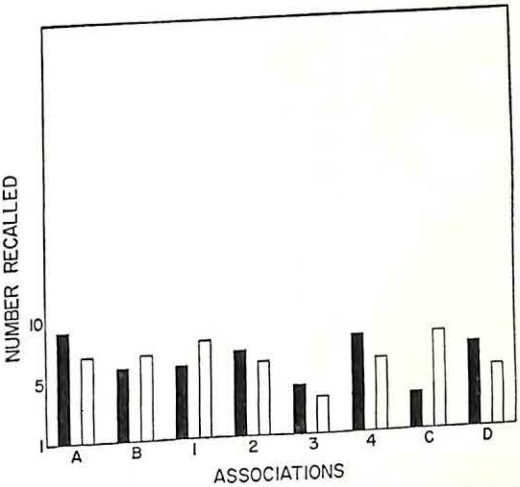


Fig. 2. Number of recalled associations 5 days after exposure to the ring of punshment. (Numbered black bars refer to punished associations and numbered white bars to control associations. Lettered black bars are rewarded associations preceding and following the punishment ring and white bars the control associations.)

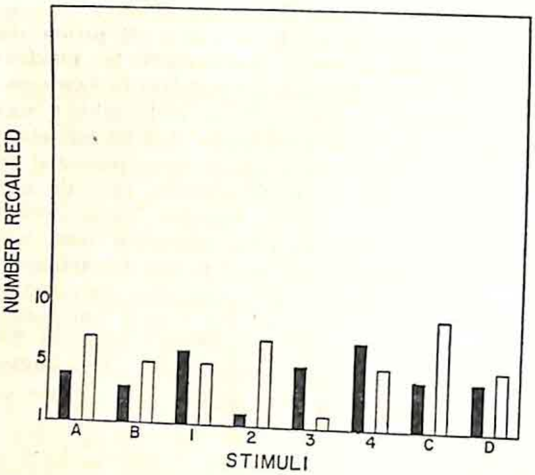


FIG. 3. Number of recalled stimulus nonsense syllables 5 days after exposure to the ring of punishment. (Numbered black bars refer to punished stimuli and numbered white bars to control stimuli. Lettered black bars are rewarded stimuli preceding and following the punishment ring and white bars the control stimuli.)

Syllables 7, 8, 9, and 10 in List 3, his recall here is compared to that on the same numbered syllables in Lists 2 and 4, respectively. In addition, since our interest was in determining possible generalization effects of punishment, we also assessed performance for the 2 syllables preceding and the 2 following the punishment ring. In all then, each subject's performance in relation to 32 syllables was evaluated.

Figures 2 and 3 show that the overall free recall of both previous associations and nonsense syllables after a 5-day interval was meager. There was only 8.7% recall of associations and 6.8% recovery of the nonsense syllables. Moreover, Figure 2 also reveals that the recall of shocked and nonshocked free associations forms virtually a straight horizontal line. Statistical analysis indicated that at no point is the difference between shock and control conditions significant. In relation to the nonsense syllables, although one may note (Figure 3) some tendencies for differences to appear between shock and control stimuli, none of these differences is statistically significant (Friedman two-way analyses of variance). Thus, in both instances free recall is found to be unrelated to the experimental conditions.

Learning

In contrast to free recall, this part of the experiment reproduced the original punishing circumstances except that real punishment was absent. Performance was assessed here in terms of errors of anticipation, defined by the subject's giving either an incorrect or no association. Each subject was again used as his own control. Figure 4 contains all the information on errors of anticipation for punished and nonpunished associations. It should be noted that the black bars for Associations 1, 2, 3, and 4 constitute the ring of punishment, while the black bars for Associations A and B and C and D are for the two associations, respectively preceding and following the punishment ring. A variety of analyses were performed. First, to test whether there was an overall difference in recall among associations, a two-way Friedman analysis of variance was performed. A χ^2_r of 55.70 (15 *df*) was found, which is significant at well beyond the .001 level. Then, two Friedman subanalyses were performed. In relation to the ring of punishment, the four shocked associations plus the two preceding and following associations provided a χ^2_r of 21.57 (7 *df*), which is significant at beyond the .001 level. A comparable analy-

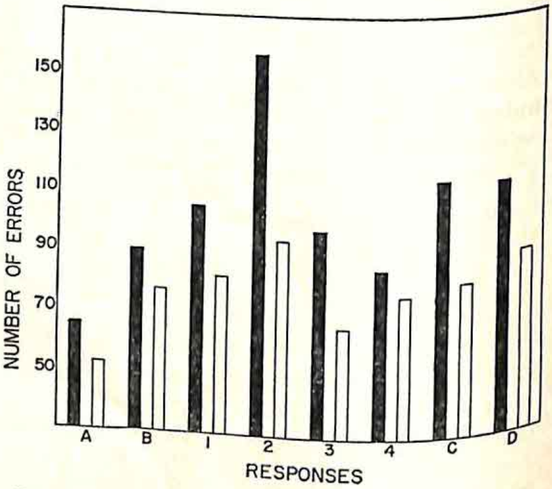


FIG. 4. Number of errors of anticipation for responses within the punishment ring and adjacent responses. (Numbered black bars indicate punished responses and numbered white bars control responses. Lettered black bars refer to rewarded responses preceding and following the punishment ring; white bars to control responses.)

sis among the eight control associations gave a nonsignificant χ^2_r of 1.28. These analyses indicate that the punishment condition was exerting a differential effect on recall (or rate of strengthening) of associations. Narrowing down to the four shocked versus the four nonshocked control associations, a two-tailed Wilcoxon test was significant at less than the .002 level. This further supports the view that punishment generally inhibited recall or the rate of strengthening of associations.

Turning to performance within the punishment ring, Figure 4 clearly shows that the second punished association had the most errors while the fourth association had the least. Our analyses indicated that this second association differed from all others within the ring at better than the 1% level and from all comparable control associations at better than the .001 level (two-tailed Wilcoxon tests). Of additional interest is the finding of a nonsignificant difference between the fourth punished association and its control which means that punishment was not exerting a net effect here. In general then, the findings within the punishment ring approximate very closely the theoretical expectations shown in Figure 1. More specifically, R_2 which is surrounded by a ring of punishment is least accessible. R_1 and R_3, which are affected by the generalization of approach gradient and the generalization from R_4, are next in the order of accessibility. Finally, R_4 which is both punished escape behavior and influenced, by the generalization from the response that was rewarded immediately afterward, is most accessible.

In connection with the foregoing findings, the question may be raised as to the possibility that shock evoked uncommon associations. However, this interpretation is not applicable since the fourth punished association should be at least and probably more uncommon than previous punished associations because the fourth association was *always* preceded by the most punishments. But our results indicate that the fourth association was most readily learned of all the punished associations and, moreover, was not significantly different from its control nonshocked association.

In addition, and although unnecessary in view of the foregoing within-subject comparisons, a frequency count was made of the associations given to each syllable. It will be recalled that 36 associations were given to each syllable. By considering as frequent those associations which appeared three or more times, comparisons were made between the shocked associations and the comparable nonshocked associations to the same syllables. Table 2 contains the number of high-frequency associations given to the four stimuli within the punishment ring and the control associations. Apart from the fact that statistical analysis is not required to determine that no reliable differences are present between the shock and nonshock conditions, it is of interest to note that there is a slight tendency for the shock conditions to produce more frequent associations.

TABLE 2

NUMBER OF HIGH-FREQUENCY ASSOCIATIONS GIVEN TO SYLLABLES WITHIN THE RING OF PUNISHMENT BY SHOCKED AND NONSHOCKED CONTROL SUBJECTS

Blocks of eight syllables	Shocked	Nonshocked
S_1	25	23
S_2 repression	23	25
S_3	30	26
S_4 suppression	25	22
Total	103	96

Note.—High frequency responses were given 3 or more times out of a total of 36 responses to each stimulus

Although within the ring of punishment the findings accord with the theory, when we move out of the ring, the picture is a bit different. Here the two rewarded associations before the punishment ring showed a trend toward more errors when compared to the control associations, but this trend only approaches significance ($p = .06$, two-tailed Wilcoxon test). More striking, however, are the error totals for the two rewarded responses following the punishment ring. Not only is there an error elevation relative to three of the responses within the punishment ring, but there is also a significantly greater number of errors ($p = < .001$, two-tailed Wilcoxon test) for the two responses combined over that for the combined comparable controls. This finding was not expected.

Effects of Number and Intensity of Punishment

Since the number and intensity of punishments received by the subjects varied widely, additional analyses were performed to determine the effects of these conditions on the production of errors of anticipation within the punishment ring. First, the total number of shocks for individual subjects ranged from 14 to 56, while for a punishment sequence they went from a low of 7 to 36. In relation to subjects, there was a trend for those who received fewer punishments (23 or less shocks) to show higher error scores (182 versus 154) than those who took more punishments (26 or more), but this was not significant. With respect to a particular punishment sequence, the same nonsignificant findings were obtained when those with 14 or more shocks (134 errors) were compared to those with 9 or fewer shocks (123 errors).

In assessing the influence of punishment intensity, those subjects who were administered 3.5 milliamperes or more were compared to those who received 2.0 milliamperes or less. The respective difference in errors (180 versus 175) is trivial. Thus, neither the number nor the intensity of punishments affected the accessibility of behaviors; the principal determinants of behavior inaccessibility were the ring of punishment and the location of behavior within this ring.

DISCUSSION

We have conceptualized repression-suppression as lying on a continuum of the accessibility of punished behavior. This accessibility of behavior was further conceived as dependent upon the location of behavior within a "ring of punishment." We have found that the least accessible behavior is that which is punished and "ringed" by other punished behaviors; hence, it is in the direction of being "repressed." In contrast, when the situation required it, the most accessible behavior was that which was *also* punished but which led to the termination of a punishment ring; hence, it is in the direction of being "suppressed."

Additional interesting properties of repression-suppression are further revealed in the absence of significant findings for the free-recall portion of the experiment where neither the independent recovery of the stimulus nonsense syllables nor the response associations were differentially affected by the ring of punishment. Instead, the punishment ring was effective in modifying behavior only when the original sequential stimulus-response situation was closely reproduced. This suggests that it is probably inaccurate to describe most people as "generally" repressed with respect to sex, dependency, etc. Thus, for example, a person might use or think a "dirty" word under some circumstances, but avoid (repress) it under those where initial sequential punishment was experienced. The importance of the stimulus situation has, of course, been frequently revealed in the somewhat related research on so-called perceptual defense. However, it is worth noting that generalized repression could occur, especially in humans, on the basis of the generation of internal stimuli which could mediate the similarity among disparate aversive situations.

Although the empirical events within the ring of punishment accord with the theoretical model, the behavior of our subjects immediately following the ring clearly does not. Here we found a greater production of errors than for several of the stimulus-response pairs within the punishment ring itself. This is indeed puzzling and no ready explanation is apparent. Some might argue from either a common sense or expectancy position that the subjects, having weathered a relatively severe sequence of experiences and, further, not knowing when punishment would be reintroduced, held a strong expectancy that they would be punished again. But this makes no sense since they were not in fact punished. Still others might apply a somewhat related notion of the generalization of punishment. This, too, will not do since the last response in the punishment ring (R_4) was associated, both theoretically and empirically, with the least avoidance. Thus, the strongest generalization should occur from this response to the postpunishment ring responses. Although some sort of generalization explanation could be invoked, our impression is that an adequate detailed account is yet to be made.

Apart from the above, and also contrary to sensible expectations, are the findings that neither the number of punishments nor their intensity differentially influenced behavior. Our views with respect to the role of intensity, however, should be tempered at the present time. The reason is that each subject here was punished with an intensity level at his upper tolerance limit. Therefore, future work will need to use intensities that are removed in varying degrees from the tolerance limit.

REFERENCES

ALEXANDER, F. *The medical value of psychoanalysis.* New York: Norton, 1932.

BREUER, J., & FREUD, S. On the psychical mechanisms of hysterial phenomena. (Orig. publ. 1893.) (Trans. by Joan Riviere) In *Collected Papers.* Vol. 1. London: Hogarth Press, 1924.

CHURCH, R. M. The varied effects of punishment on behavior. *Psychological Review,* 1963, 70, 369–402.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

FREUD, S. Psycho-Analytic notes upon an autobiographical account of a case of paranoia. Dementia Paranoides (Orig. publ. 1911) (Trans. by Joan Riviere) In, *Collected papers.* Vol. 3. London: Hogarth Press, 1925. (a)

FREUD, S. Repression. (Orig. publ. 1915) (Trans. by Joan Riviere) In, *Collected papers.* Vol. 4. London: Hogarth Press, 1925. (b)

GLAZE, J. A. The association value of nonsense syllables. *Journal of Genetic Psychology,* 1928, 35, 255–269.

GRINKER, R. R., & SPIEGEL, J. P. *Men under stress.* Philadelphia: Blakiston, 1945.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *Journal of Experimental Psychology,* 1934, **17**, 145–153.

MACKINNON, D. W., & DUKES, W. F. Repression. In L. Postman (Ed.), *Psychology in the making.* New York: Alfred A. Knopf, 1962.

POSTMAN, L. Rewards and Punishments in Human Learning. In L. Postman (Ed.), *Psychology in the making.* New York: Alfred A. Knopf, 1962.

SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Social Science Research Council Bulletin,* 1943, No. 51, 1–156.

(Received April 10, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 210–213

SELF-REGARD AS A FUNCTION OF ENVIRONMENTAL DEMANDS IN CHRONIC SCHIZOPHRENICS[1]

GUSTAVE MANASSE[2]

Teachers College, Columbia University

This study explored the effect of the social setting upon the self-regard of chronic schizophrenics. It was hypothesized that self-regard is related to the degree to which a person is able to meet the demands and expectations of his social setting. 2 groups of comparable chronic schizophrenics were compared. Group 1 was hospitalized and Group 2 attended a day treatment center. Hilden's Q sort procedure was used to obtain a measure of self-regard. Results revealed that the hospitalized group had higher self-regard than the nonhospitalized group. The findings were interpreted as shedding more light on the importance of the situational variable in the development and maintenance of self-regard.

In recent years many sociologists and social psychologists have stressed the importance of the social setting upon the behavior and thinking of individuals. Studies of the psychiatric hospital particularly have emphasized that the structural properties of the social organization are an important determinant of behavior. Henry (1957) states that much human behavior occurs as a direct consequence of the structural properties of the social organization within which a person finds himself. Parsons (1957) views the hospital as a unique type of social organization very different from the community. Stanton and Schwartz (1954) point out that the hospital removes the patient from the increasingly interpersonal situations and complexities of the community and exposes him to patients who act very much as he does and staff who encourage his patient role. It is clear that when psychiatric patients are removed from society they are also removed from its demands and expectations and are placed in a setting that is far less complex and far less demanding.

Studies comparing the self-regard of psychopathological subjects with that of other psychopathological subjects, of medical patients, and of normal subjects are numerous in the literature, but none were found in which the important variable of social setting had been taken into account. Thus, it is not possible to tell whether the results obtained have general meaning and applicability or whether they have meaning only for the groups in the settings involved. Despite the great research interest in this area, results have often been conflicting, possibly for the reasons stated above. While some researchers (Chase, 1957; Tamkin, 1957; Tolor, 1957; Zuckerman, Baer, & Monashkin, 1956) found that psychiatric patients had lower self-regard than normal subjects, others (Epstein, 1955; Friedman, 1955; Hillson & Worchel, 1957) found either no difference in self-regard or higher self-regard among psychiatric patients as compared to normals. It may well be that the conflicting results are due, at least in part, to the effects of the different environmental expectations. The kind of adjustment the studies deal with may be adjustment within a specific setting rather than adjustment in a more general sense.

Studies of the psychiatric hospital (Greenblatt, Levinson, & Williams, 1957; Stanton & Schwartz, 1954) emphasize that the difference between the hospital and the community extends not only to social organization but also to the different expectations and demands that each makes upon the individual. All researchers are agreed that the demands and expectations of the hospital environment are far less than the demands and expectations of the community in so far as optimal adjustment within the settings is concerned. The hospital setting demands a smaller repertoire of responses for a satisfactory adjustment than does the community. But as recently as 1961 Crowne and Stephens still noted that not much is known about the influence of situational variables in determining self-regard.

---

[1] Based on the author's doctoral dissertation at Teachers College, Columbia University, under the sponsorship of Donald E. Super and Rosalea A. Schonbar.

[2] Now at Children's Rehabilitation Service, Long Island Jewish Hospital, Queens Hospital Center Affiliation, New York.

The general purpose of the present study is to investigate the effect of differing environmental expectations and demands upon self-regard. Specifically, the study investigates two groups of chronic schizophrenics, comparable except that the members of one group are hospitalized because no family is available to care for them in the community, whereas the members of the other group are not hospitalized, largely because they have families available to care for them. In addition, the study also seeks to shed more light on the assumption that positive self-regard is an index of general adjustment (Rogers & Dymond, 1954) rather than of adjustment to a specific setting which is based on fulfillment of the demands and expectations of that setting.

The study is concerned with the actual-self-concept and ideal-self-concept discrepancy, as this discrepancy is related to the subject's perceptions of environmental demands and expectations in his social setting.

In attempting to verify the general hypothesis that environmental demands and expectations affect self-regard, the following specific hypotheses are examined: (*a*) Hospitalized chronic schizophrenics have a higher correlation between actual self-concept and ideal self-concept than do nonhospitalized schizophrenics. (*b*) Hospitalized chronic schizophrenics have a higher correlation between actual self-concept and how they feel the environment expects them to be than do nonhospitalized chronic schizophrenics.

METHOD

Subjects

Two groups of male chronic schizophrenics were used in the study. Group 1 consists of 51 patients in a Veterans Administration Neuropsychiatric Hospital. Group 2 is made up of 51 patients reporting regularly for treatment at a Veterans Administration Neuropsychiatric Day Center.

The Hospital Group was composed of patients in the "best" chronic wards. The principal reason for the continued hospitalization was that no family care was presently available for them in the community. Patients in this group enjoyed all the privileges of the hospital and all had work or other assignments within the hospital for the better part of the day. Were these patients to be discharged, they would be the potential patients of the Day Center.

The Day Center Group was composed of patients in a Veterans Administration Neuropsychiatric Day Center. The Day Center was part of the Mental Hygiene Clinic and its patient population was composed of former neuropsychiatric hospital patients discharged to their families but unable to hold remunerative employment or independently to manage their own day time activities. The Day Center provided a sheltered setting in which patients engaged in various activities of a recreational, educational, and prevocational nature.

Subjects were selected by staff psychologist to meet the following criteria: (*a*) a minimum of 3 years continuous hospitalization in a neuropsychiatric facility, (*b*) diagnosis of chronic schizophrenia in partial remission, (*c*) at least a sixth grade education, (*d*) at least 6 months in the Day Center (Group 2), (*e*) ready for discharge but lacking an accepting family or suitable residence (Group 1). One hundred and two subjects were used, 51 in each group.

After subjects were selected, clinical folders were examined to obtain data with respect to such variables as age, education, religion, marital status, and total years of hospital treatment. These data were examined to determine whether the two groups were comparable on these variables. Chi-square analyses of the results indicated that the groups did not differ significantly. Social service records of hospital patients were examined in order to determine whether these subjects were, in fact, kept in the hospital because discharge planning could not proceed normally. It was thus possible to confirm whether discharge planning had been initiated at some time prior to the subject's selection for the experiment, when it had been initiated, and why it had been delayed or discontinued.

Examination of the records indicated that in every case discharge planning was initiated some time before the experimenter tested the subjects. The records revealed that in some cases discharge planning had begun as long as 3 years prior to the study.

Procedure

Set Number 13 of the Q-sort procedure devised by Hilden (1954, 1958) based on a random set of personal concepts was used to obtain an actual-self sort, an expected-self sort, and an ideal-self sort. The same order of presentation was used for all subjects. An attempt was made to minimize carry-over and response set by administering a questionnaire used for other research between the expected-self sort and the ideal-self sort.

Treatment of data

The basic statistical methods employed in this study were median tests and chi-square. Yates' correction for small samples was applied to all chi-square calculations. The .01 level was set for accepting the hypotheses.

Results and Discussion

The prediction that self-regard was related to environmental demands and expectations was borne out. Hypotheses 1 and 2, both of which attempted to establish that subjects' self-regard was related to the environmental demands and expectations of their respective social settings, were confirmed. The differences in self-regard between the two groups, as demonstrated by Table 1, were significant at the .001 level.

The findings suggest that the specific environment in which a person finds himself influences his thinking about himself. Different environments create different standards by which people evaluate themselves. An environment that demands little of the marginal individual will enable him to maintain a relatively positive self-image; on the other hand, an environment that demands a great deal of the same individual will lead to a relatively negative self-image. This is clearly demonstrated in the two groups of chronic schizophrenics studied here. The Hospital Group which is considered the best adjusted in the patient population enjoying maximum privileges and responsibility, can be expected to and does look upon itself in a relatively favorable light. Since the hospital clearly demands less of the individual than does the community, it does not constantly remind him of his inadequacies, especially if he has been there for some time. It therefore enables him to create a new life for himself within the setting and to reject or deny the existence of the external world. He can create for himself the illusion of doing very well in general while really doing well only in the hospital setting. On the other hand, subjects in the Day Center, who are equally "disabled" but who live in an environment that demands greater achievements from its members, are unable to maintain high self-regard. They, unlike the Hospital Group, are reminded daily of their inability to function by their exposure to the external world and hence cannot maintain positive self-regard without distorting reality.

TABLE 1

Actual-Ideal and Actual-Expected Self-Concept Median Correlations of Hospital (I), Day Center (II) and Combined Groups of Chronic Schizophrenics

Type correlation	Median Group 1 $N = 51$	Median Group 2 $N = 51$	Median combined $N = 102$	Chi-square with Yates correction for median test
Actual-ideal	.73	.24	.60	90.3*
Actual-expected	.67	.34	.54	62.7*

* $p \leq .001$.

It is quite possible that the conflicting result obtained in previous studies evaluating the self-regard of pathological subjects was failure to take into account the effects of varying social settings.

Since high self-regard appears to be characteristic of the Hospital Group and low self-regard characteristic of the Day Center Group, the question that must be asked is: "What happens at the time of discharge?" The artificially high self-regard engendered in the hospital probably makes it difficult for patients to make an effective transition to the community. The relative artificiality and simplicity of the hospital setting, the absence of the tensions and anxieties of the external environment, permits patients to experience a rise in self-regard and instills them with a false sense of security. As they leave the sheltered setting of the hospital and return to the community they find that their view of themselves and of their performance needs quick and drastic reappraisal. The degree of self-regard which was appropriate in the hospital is no longer appropriate in the community. Instead of being able to feel that they are doing well they find that, when compared with other people in the community, they function very poorly. Thus, the rapid shift from high self-regard to lowered self-regard may contribute to their rehospitalization.

Reducing the isolation of psychiatric patients from their community might be one way of preventing the building of an artificially high self-regard, the maintenance of which creates adjustive difficulties. A program of progressively increased expectations and demands might help patients to effect the transition back to the community without loss of self-regard and consequent trauma.

Awareness of the difficulties faced by the long-term hospital patients in adjusting to the outside community should lead not only to better preparation within the hospital by instituting a program of progressively increased expectations, but also to the formation of more transitional facilities and out-patient services in which support is provided during the attempt at reintegration in the outside world. The incidence of rehospitalizations of psychiatric patients may well be due to lack of preparation before leaving the hospital and to the lack of transitional facilities within the community, leaving patients dependent on their own limited and usually inadequate resources.

REFERENCES

CHASE, P. H. Self-concept in adjusted and maladjusted hospital patients. *Journal of Consulting Psychology,* 1957, **21**, 495–497.

CROWNE, D. P., & STEPHENS, M. W. Self-acceptance and self-evaluative behavior: A critique of methodology. *Psychological Bulletin,* 1961, **58**, 104–121.

EPSTEIN, S. Unconscious self evaluation in a normal and a schizophrenic group. *Journal of Abnormal and Social Psychology,* 1955, **50**, 65–70.

FRIEDMAN, I. Phenomenal, ideal and projected conceptions of self. *Journal of Abnormal and Social Psychology,* 1955, **21**, 243–247.

GREENBLATT, M., LEVINSON, D. J., & WILLIAMS, R. H. *The patient and the mental hospital.* Glencoe, Ill.: Free Press, 1957.

HENRY, J. Types of institution structure. In M. Greenblatt, D. J. Levinson, & R. H. Williams (Eds.), *The patient and the mental hospital.* Glencoe, Ill.: Free Press, 1957. Pp. 73–90.

HILDEN, A. H. *Manual for Q-Sort and random sets of personal concepts.* St. Louis, Author, 1954. (Mimeo)

HILDEN, A. H. Q-Sort correlation: Stability and random choice statements. *Journal of Consulting Psychology,* 1958, **22**, 45–50.

HILLSON, J. S., & WORCHEL, P. Self-concept and defensive behavior in the maladjusted. *Journal of Consulting Psychology,* 1957, **21**, 83–88.

PARSONS, T. The mental hospital as a type of organization. In M. Greenblatt, D. J. Levinson, & R. H. Williams (Eds.), *The patient and the mental hospital.* Glencoe, Ill.: Free Press, 1957. Pp. 108–129.

ROGERS, C. R., & DYMOND, R. (Eds.), *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

STANTON, A. H., & SCHWARTZ, M. S. *The mental hospital.* New York: Basic Books, 1954.

TAMKIN, A. S. Selective recall in schizophrenia and its relation to ego strength. *Journal of Abnormal and Social Psychology,* 1957, **55**, 345–349.

TOLOR, A. Self perceptions of neuropsychiatric patients on the W-A-Y tests. *Journal of Clinical Psychology,* 1957, **13**, 403–406.

ZUCKERMAN, M., BAER, M., & MONASHKIN, I. Acceptance of self, parents and people in patients and normals. *Journal of Clinical Psychology,* 1956, **12**, 315–320.

(Received May 4, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 214–217

SOCIAL AND NONSOCIAL VERBAL CONCEPTS OF DEAF AND HEARING CHILDREN[1]

SOLIS L. KATES AND FRANCES F. KATES[2]

University of Massachusetts and Clarke School for the Deaf, Northampton

Social and nonsocial verbal concepts of 24 deaf children were compared with those of 24 hearing children matched on IQ and age, and with those of 24 hearing children matched on IQ and school achievement. The deaf Ss gave the same number of synonym definitions for social and nonsocial words and had more errors for social than for nonsocial words. Each hearing group gave significantly more synonyms for social than for nonsocial words and had the same number of errors. The deaf and achievement control Ss were not significantly different on the number of synonyms and errors offered for all words. The age control group gave more synonyms and fewer use-description definitions than the deaf Ss.

The purpose of this study was to compare social and nonsocial verbal concepts of deaf children with those of hearing children matched on IQ and age (age controls), and with those of hearing children matched on IQ and school achievement (achievement controls). In this study, words are considered as verbal concepts inasmuch as each word refers to a class of objects and events (Vygotsky, 1962).

The primary source of language for the deaf child is the school where he is formally taught words and their meanings. Therefore, it is anticipated that there would be no difference between his understanding of social and nonsocial words. The hearing child, on the other hand, uses more social than nonsocial words to describe the attributes of his predominantly social environment. When he is given systematic instruction in school, the hearing child will reorganize his previous interpretations of these social words to arrive at more formal class definitions (synonyms). Moreover, his use-description definitions and his errors for the social words will decrease in comparison with the nonsocial words.

The deaf child finds it difficult to include the appropriate referent objects and events that belong to the category embraced by a word and to eliminate those that do not. Consequently, he is retarded in the proper use of the verbal symbol, a retardation that falls along normal developmental lines (Kates, Kates, & Michael, 1962). Hence, the deaf subjects and the achievement hearing control subjects, who are similar to the deaf subjects on school attainment, would not differ significantly in number of synonyms, use-description definitions, and errors for all words, social and nonsocial combined. However, when compared with the age controls (who are a year advanced in school achievement), the deaf subjects would be less able to analyze all words into synonyms and would give more use-description definitions and make more errors.

The hypotheses are as follows:

1. There will be a significant interaction between the deafness-hearing condition and the social-nonsocial word dimension, the two independent variables. The deaf group will perform equally well on social and nonsocial words whereas each hearing group will show superior performance on social as compared with nonsocial words.

2. There will be no significant differences between the deaf group and the achievement control group on the measures for all words, social and nonsocial combined. However, the age control group will be significantly superior to the deaf group on the measures for all words, social and nonsocial combined.

In both hypotheses, the word definitions are analyzed into measures which include synonyms, use-descriptions, and errors.

---

[1] This study was supported, in part, by Research Grant HD336 from the National Institutes of Health, United States Public Health Service.

[2] Portions of this paper were presented at the Eastern Psychological Association meetings at Philadelphia, April 17, 1964.

METHOD

Subjects

Each of the three groups, the deaf, the age controls, and the achievement controls, was composed of 24 children, 12 boys and 12 girls. The average age of the deaf children was 11 years and 9 months, *SD* 11 months, their average IQ was 95, *SD* 13, and their average school achievement grade score was 4.3, *SD* 1.2. All deaf subjects attended Clarke School, a private school for the deaf in Northampton, Massachusetts, and had hearing losses of not less than 80 decibels. Both control groups were selected from the Northampton, Massachusetts school system. The age control subjects were matched with the deaf subjects on age (11 years and 9 months, *SD* 10 months) and IQ (96, *SD* 13). The achievement controls were matched with the deaf group on Stanford Achievement Test marks (4.3 years, *SD* 1.2) and IQ (96, *SD* 11). The groups were equated on socioeconomic background as determined by the fathers' occupations. The distribution of these occupations among professional and semiprofessional, skilled, semiskilled, clerical, and manual workers was the same for all three groups.

Test Materials

Sixteen words were selected from the 500 most frequently used words as listed in the Thorndike-Lorge Word Book. Two teachers of deaf children agreed that these 16 words should be known by deaf children whose age and educational level were comparable to our deaf subjects.

These 16 words were rated on a 5-point scale by 12 judges, all psychologists (including graduate students), on the degree to which these words described or involved social relationships. A rating of one characterized the greatest degree of social relationship and a rating of five described the lowest degree of social relationship. The 8 words, rated as most characteristic of social relationships and known as the social words, averaged 1.8 (family, fear, friend, love, smile, home, man, and woman). The 8 words, rated as least characteristic of social relationships and known as the nonsocial words, averaged 3.9 (color, world, sun, name, heart, voice, question, and doctor).

Procedure

The 16 words were printed in random order in a test booklet with space provided for writing the definition of each word. The test booklets were distributed to the subjects with the following instructions: "I want to see how many words you know. Write what each word means." Sufficient time was made available for each subject to write as much as he wished about each word. The scoring system, adapted from Feifel (1949), included three categories: synonyms, used-description, and errors. These categories formed the dependent variables. Two scorers who independently evaluated the responses agreed on 94% of the 320 items scored.

RESULTS

The First Hypothesis

Synonym measure. There was a significant interaction between the condition of deafness-hearing and the social-nonsocial word dimension on the synonym variable (Table 1). The deaf subjects offered the same number of synonyms for the social words as for the nonsocial words (Table 2). In contrast, the achievement control group and the age control group each offered significantly more synonyms for social than for nonsocial words (Table 2).

TABLE 1

ANALYSIS OF VARIANCE FOR DEPENDENT VARIABLES

Source	Synonym			Use-Description		Error	
	df	*SS*	*F*	*SS*	*F*	*SS*	*F*
Between	71	492		244		343	
Groups (A)	2	88	7.4**	31	5.1*	22	2.2
Sex (B)	1	4	.7	4	1.2	0.0	0
A × B	2	9	.7	9	1.5	0.3	0
Ss/AB	66	391		201		321	
Within	72	138		159		75	
Words (C)	1	25	17.6***	41	24.3***	2	2.1
A × C	2	18	6.3**	0.5	.2	15	8.8***
B × C	1	1	.7	0.5	.3	0	.0
A × B × C	2	1	1	4	1.4	3	1.6
Ss × C/AB	66	93		112		55	
Total	143	630		403		418	

* $p \leq .01$.
** $p \leq .005$.
*** $p \leq .001$.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR DEPENDENT VARIABLES

	Synonym		Use-Description		Error	
	M	*SD*	*M*	*SD*	*M*	*SD*
Deaf						
Social	1.67	2.04	3.47	1.80	2.83	1.90
Nonsocial	1.79	1.87	4.54	1.84	1.71	1.72
Total	1.73	1.96	4.00	1.82	2.27	1.81
Hearing (Age)						
Social	4.42	2.12	2.34	1.48	1.29	1.46
Nonsocial	2.88	1.82	3.54	1.24	1.59	1.31
Total	3.65	1.98	2.94	1.37	1.44	1.39
Hearing (Achieve)						
Social	3.17	1.88	2.67	1.37	2.17	2.03
Nonsocial	2.09	1.73	3.59	1.40	2.33	1.62
Total	2.63	1.81	3.13	1.39	2.25	1.84

Use-description measure. There was no significant interaction between the condition of deafness-hearing and the social-nonsocial word dimension on the use-description measure (Table 1). Each group gave more use-description definitions for nonsocial than for social words (Table 2).

Error measure. There was a significant interaction between the condition of deafness-hearing and the social-nonsocial word dimension on the error measure (see Table 1). The deaf group made significantly more errors on social than on nonsocial words (see Table 2). The achievement control and the age control groups did not differ significantly in the number of errors made on social and nonsocial words.

The Second Hypothesis

Synonym measure. As indicated by Table 1 and by Duncan multiple-range tests, the deaf and the achievement control subjects did not differ significantly on synonyms, but the age control subjects gave a significantly greater number of synonyms than the deaf subjects for all words.

Use-description measure. For all words, the deaf and hearing subjects differed significantly on the use-description measure (see Table 1). Further analyses by the Duncan test disclosed that the deaf subjects gave significantly more use-description definitions for all words than either the age or the achievement control subjects (see Table 2).

Error measures. For all words, the three groups did not differ significantly on errors (see Tables 1 and 2).

Additional Statistical Analyses

Additional analyses were carried out, comparing the three groups first on social and then on nonsocial words.

Synonym measure. As indicated by Table 3 and by the Duncan test, the deaf group offered significantly fewer synonyms for social words than the two hearing groups. However, the three groups did not differ significantly on nonsocial words (see Table 3).

Error measure. The deaf subjects made significantly more errors on social words than the age control subjects but did not differ in this regard from the achievement controls. No significant differences in errors appeared for nonsocial words.

Finally, males and females did not differ significantly in any of the analyses.

DISCUSSION

The First Hypothesis

The hearing groups offered significantly more synonyms for social than for nonsocial words, but the deaf offered about the same number of synonyms for both. Although the remainder of this hypothesis relating to use-description definitions and errors was not supported, it should be recalled that the deaf

TABLE 3

ANALYSIS OF VARIANCE FOR THE SOCIAL WORDS AND FOR THE NONSOCIAL WORDS

Words	Synonyms			Use-Description		Errors	
	df	*SS*	*F*	*SS*	*F*	*SS*	*F*
Social							
Groups	2	91	11.7**	16	2.1	29	4.5*
Error	69	273		263		218	
Total	71	364		279		247	
Nonsocial							
Groups	2	16	2.4	15	3.0	7	1.0
Error	69	226		168		162	
Total	71	242		183		169	

* $p \leq .025$.
** $p \leq .001$.

subjects made significantly more errors on social than on nonsocial words, whereas both hearing groups did not differ in errors. These findings clearly indicate that deaf subjects do not understand social words as well as they understand nonsocial words. In contrast, the hearing subjects understand social words better than nonsocial words. Our results from the additional statistical analyses further support this conclusion.

Possibly, social experiences require more verbalizations than do nonsocial referents to be readily understood. In this connection, the hearing would have had more verbal communications and evaluations about their understanding of social experiences and therefore greater opportunity to develop many different synonyms for these social experiences (Brown, 1958; Whorf, 1956). Without the equivalent opportunity for such verbalizations, the deaf would do no better in synonyms for social than for nonsocial words.

The deaf make more errors on social words because these words have more elusive and transitory properties. Hence, social words are more difficult to understand than nonsocial words. Due to their hearing impairment, the deaf children tend to persist in these errors inasmuch as feedback and correction are more difficult for them.

The Second Hypothesis

The deaf and achievement control groups were not significantly different on the number of synonyms and errors, thus confirming this hypothesis in part. In their performance on synonyms, the two groups were equally able to verbalize an acceptable synonym, or a phrase conveying the main idea, or an appropriate technical definition. Apparently, deaf children and those hearing children who are matched on school achievement show comparable skill in attaching the appropriate words to categories of nonlinguistic reality, and in having these words signify the appropriate properties of these categories.

The error-measure finding demonstrated that the deaf and achievement control subjects do not differ in idiosyncratic-personal definitions. Both groups offer a similar number of misconceptions or vague definitions for all words. Further, the deaf subjects gave more use-description definitions than the achievement control subjects. Method of teaching word meanings to the deaf may be partly responsible for this difference. It is certainly simpler to point to some perceptible characteristic or use of some word than to represent it by a synonym or technical definition.

The age control subjects gave more synonyms and fewer use-description definitions than the deaf subjects. Possibly, the advanced school achievement of the age controls contributed to better performance. These results provide some evidence that the use of synonyms develops from a knowledge of functional and perceptible attributes of the word's referents. To support this inference, it should be noted that the deaf subjects do not differ significantly from the age controls on errors, but give fewer synonyms and more use-description definitions. If their language retardation lies along normal developmental lines, then deaf children will transform their functional-perceptible definitions into synonyms when they acquire more education and experience.

REFERENCES

Brown, R. *Words and things.* Glencoe, Ill.: Free Press, 1958.

Feifel, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genetic Psychology Monographs,* 1949, **39**, 151–204.

Kates, S. L., Kates, W. W., & Michael, J. Cognitive processes in deaf and hearing adolescents and adults. *Psychology Monographs,* 1962, **76**(32, Whole No. 551).

Vygotsky, L. S. *Thought and language.* Cambridge, Mass.: M.I.T. Press, 1962.

Whorf, B. L. *Language, thought, and reality.* New York: Wiley, 1956.

(Received May 6, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 218–220

VOICE QUALITY IN DEPRESSION[1]

WILLIAM A. HARGREAVES, J. A. STARKWEATHER, AND K. H. BLACKER

Langley Porter Neuropsychiatric Institute and The University of California School of Medicine

32 depressed patients were interviewed daily throughout their hospitalization. Spectrum analyses of the patients' speech were compared with mood ratings made by the 2 clinicians who jointly conducted the interviews. Using a stepwise multiple regression with cross-validation for each patient, predictions of the mood ratings from the voice spectra correlated significantly (1% level) with the ratings for 25 of the 32 patients. These cross-validation correlations were largest for those patients who showed considerable change in their mood.

Severe depression is accompanied by a variety of behavioral changes. The clinician may see changes in patterns of movement and speech, in postural and facial indications of apathy or anxiety, and changes in the thought content reported by the person. While there has been recent interest in developing systematic rating techniques to assess depression (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Grinker, Miller, Sabshin, Nunn, & Nunnally, 1961; Hamilton & White, 1959; Hamilton, 1960; Overall, 1962; Wechsler, Grosser, & Busfield, 1963), very little has been attempted in the way of direct measurement of the behavior of depressed patients. Furthermore, most of the attention in current studies of depressive symptomatology is given to differentiating groups of patients, while relatively little attention is given to detailed measurement of change in patients over time.

We have explored a method of measuring voice quality, the sound of the voice rather than the content. Changes in voice quality frequently accompany depression, and measurement of these changes is a first step toward the comprehensive measurement of depressive symptoms. The voice changes, as with other behavioral changes in depression, are easily observable, but not so easy for the observer to record reliably, or remember well enough to make accurate comparisons over long periods of time. Direct measurement of these symptoms will provide potentially valuable criterion measures in studies of the treatment of depression.

METHOD

The subjects were 32 depressed patients hospitalized on the Acute Treatment Service at the Langley Porter Neuropsychiatric Institute. The group consisted of 23 women and 9 men, ranging in age from 20 to 72. Their hospital stay was typically about 5 weeks. A psychiatrist (KHB) and psychologist (WAH) jointly interviewed the patients each day they were in the hospital. The interviews lasted 5 to 10 minutes. The interview began with a question about how the patient felt that day, and after 4 or 5 minutes of inquiry about feelings, concluded with a few standard questions aimed at the assessment of confusion and memory loss, since most of these patients were receiving some kind of somatic therapy.

The interviews were recorded with an Electrovoice 649A lavalier microphone hung about the patient's neck and throat microphones for the two interviewers. The combined output of the two throat microphones was recorded on a second channel of the tape, providing a clear indication of when anyone but the patient was speaking. At the end of each session the two interviewers rated the patient on an inventory designed for the study, which included items related to mood and to confusion and memory loss. Individual items were rated on a scale from zero ("not at all") to eight ("extremely"), although the raters extended the scale when it seemed required by the patient's behavior. The original rating inventory was revised after the first 13 patients to include many items from the Brief Psychiatric Rating Scale (Overall & Gorham, 1962). From each version of the inventory, a group of items was combined to yield an overall mood rating for each interview. In the initial version of the inventory the items were rationally selected, while for the revised inventory the item selection was based on a factor analysis of the items. For this item analysis, the third, middle, and third from last interview ratings were selected from each of the 19 subjects. Scores from the two raters were combined. A principal component analysis extracted six factors which were rotated by both the varimax and quartimax methods. The two rotations yielded virtually identical results, indicating a group of items which could form an internally homogeneous scale of mood. These were all items taken from the Brief Psychiatric Rating Scale. Listed in the order of rotated factor loadings, they were the items "depressive mood," "anxiety," "motor retardation," "tension," and "guilt feelings." For each of the 32 subjects the scores on the mood items were summed for each rater separately, and the interrater agreement across interviews was calculated. This reliability ranged from .51 to .99, except for one patient who showed little mood change, on whom agreement dropped to .34. The median interrater correlation was .84.

---

[1] Presented at Western Psychological Association, Portland, April 1964. This investigation was supported by Research Grant MH-03375 from the National Institute of Mental Health, United States Public Health Service, and additionally assisted by Public Health Service Research Grant FR-00122.

In analyzing the recordings our task was to discover what, if any, measurable changes occurred in the sound of the patient's voice when he was depressed. Some obvious things to look for are whether the patient talks loudly or softly, with a high voice or a low voice, with a smooth voice or a harsh voice. Characteristics like these are reflected in the power spectrum of the voice. The voice can be thought of as a mixture of tones of different pitch, with a fundamental pitch at the low end and a series of higher overtones. Just as the timbre of musical instruments corresponds to the way in which various overtones are emphasized, so does voice quality depend upon which overtones in the voice are emphasized. The voice spectrum is a direct measure of the relative prominence in the voice of the fundamental pitch and the various higher overtones. The spectrometer used in this study (Hargreaves & Starkweather, 1963) receives the recorded sound of the patient's voice from the tape-recorder playback, and produces a set of numbers on punch cards representing the voice spectrum. Editing in the computer then produces an average spectrum for each consecutive 5 seconds of the patient's speech.

RESULTS AND DISCUSSION

We did not expect the same kind of depressed voice quality in every subject. Therefore it was appropriate to study each patient's voice to discover his own particular pattern of change. A multiple-regression procedure was used to develop a method of predicting the mood ratings from the voice spectra of a particular patient. The regression weights were computed using only spectra from even-numbered interviews. The obtained multiple correlations, one for each patient, ranged from .57 to .97, with a median of .83. The spectra from the odd-numbered interviews were then used to cross-validate the regression weights. The mood predictions computed from these additional spectra were correlated with the mood ratings for those interviews. In 25 out of 32 patients these cross-validation correlations were significant at the .01 level or beyond. While this means that for 25 patients some aspect of the voice changed with mood, it does not mean that the mood ratings could be accurately predicated in all of these patients. Several correlations were about .80, but in other patients the relationship, while statistically significant, was too weak to be of any practical significance.

The size of the cross-validation correlation was related to how much change occurred in the patient's symptoms. That is, the more the patient's mood changed while he was in the hospital, the more clearly we could see changes in his voice. The variability of the subjects' mood ratings correlated .61 with the size of the cross-validation correlation. If mood neither improved nor worsened substantially, then it was naturally more difficult either to see changes in the voice or to obtain reliable ratings of mood change.

The voice changes found do not appear to be an artifact related to the effects of somatic treatment, though we could not experimentally control these variables. Most of these depressed patients were treated with amitriptyline (Elavil), and some also received a course of six to eight ECT. Amitriptyline dosage was as high as 250 milligrams per day for some patients, and several showed mild confusional reactions while on the drug. The ECT often produced a pronounced confusional reaction. Multiple regression of the voice spectra onto the daily ratings of confusion were computed. Eighteen of the 32 patients had significant cross-validation correlations, suggesting that this type of organic impairment may be reflected in the voice of some patients. These 18 patients did not, however, have higher mood-voice correlations than the remaining 14 patients, the mean correlation being .41 in both groups. The drug regimen of these patients was quite variable, which made it difficult to assess drug effects on the voice. Drug dose per day occasionally correlated highly with rated mood, negatively as often as positively. Therefore any drug effect on the voice could have spuriously produced a mood-voice relationship. However, patients with high drug-mood correlations did not have a greater tendency to show a significant mood-voice relationship. Furthermore, in a previous study of drug effects on the voice (Starkweather & Hargreaves, 1964) we found no clear effects of 150 milligrams of pentobarbital on the voice spectrum, even though audible slurring occurred in several subjects. While these considerations make it unlikely that drug effects account for our findings, further work in this area would benefit from systematic control of medication and ECT.

Is there anything like a typical voice quality of depression? Since the clearest picture of this could be obtained from the patients who showed large changes, we examined the kinds of voice changes which occurred in 10 patients with large cross-validation correlations. Five of these showed a depressed voice reduced in loudness and often lacking in high overtones, giving their voice a dull, lifeless quality with diminished inflection. Their quiet voice did not always drop in fundamental frequency, but sometimes rose in frequency in a kind of pathetic whimper. In addition there were 2 or 3 patients who showed a depressed voice which was actually louder during depression, and sometimes higher in pitch.

Recent factor analytic studies (Grinker et al., 1961; Hamilton, 1960; Overall, 1962) have found symptom clusters which may relate to these various types of voice change. As we become clearer about the types of change which occur, it should be possible to develop ways to measure directly some of these clusters of depressive symptoms. The more prominent symptoms should yield fairly soon to these direct measurement techniques.

REFERENCES

BECK, A. T., WARD, C. H., MENDELSON, M., MOCK, J., & ERBAUGH, J. An inventory for measuring depression. *Archives of General Psychiatry,* 1961, **4**, 561–571.

GRINKER, R. R., MILLER, J., SABSHIN, M., NUNN, R., & NUNNALLY, J. C. *The phenomena of depressions.* New York: Harper, 1961.

HAMILTON, M. A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry,* 1960, **23**, 56–62.

HAMILTON, M., & WHITE, J. M. Clinical syndromes in depressive states. *Journal of Mental Science,* 1959, **105**, 985–998.

HARGREAVES, W. A., & STARKWEATHER, J. A. Recognition of speaker identity. *Language and Speech,* 1963, **6**, 63–67.

OVERALL, J. E. Dimensions of manifest depression. *Journal of Psychiatric Research,* 1962, **1**, 239–245.

OVERALL, J. E., & GORHAM, D. R. The brief psychiatric rating scale. *Psychological Reports,* 1962, **10**, 799–812.

STARKWEATHER, J. A., & HARGREAVES, W. A. The influence of sodium pentobarbital on vocal behavior. *Journal of Abnormal and Social Psychology,* 1964, **69**, 123–126.

WECHSLER, H., GROSSER, G. H., & BUSFIELD, B. L. JR. The depression rating scale. *Archives of General Psychiatry,* 1963, **9**, 334–343.

(Received May 13, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 221–224

STUTTERING AS A FUNCTION OF THE INFORMATION VALUE AND SENTENCE POSITION OF WORDS

BRUCE QUARRINGTON

University of Toronto

24 stutterers read aloud a 95-word passage of banal prose, and the incidence of stuttering was determined for each word. The information value of each word was estimated by studying the extent to which each word could be predicted from knowledge of the preceding words by 46 normal speakers. Results indicated that both position in the sentence and information value of words are significantly correlated with stuttering incidence. It is concluded that the positional phenomenon cannot be explained as a redundancy effect. Its nature is discussed briefly from the approach-avoidance conflict and operant viewpoints of stuttering.

When stuttering is related to the position a word occupies in the sentence a declining gradient of incidence is usually obtained. Words occurring in initial or early positions are stuttered upon much more frequently than words occurring later in the sentence structure (Brown, 1938, 1945). While such trends have usually been studied in reading prose material aloud, similar trends have been found in the spontaneous speech of stutterers (Hejna, 1955).

The word-position gradient of stuttering has been explained in terms of the conspicuousness of early words and the greater burden of information conveyed by words in earlier positions (Brown, 1943, 1945). It is argued that the stutterer monitors the communicative importance of the words he is to speak and mobilizes greater avoidance of nonfluency on these words of greater informational value. It is believed that the typical stuttering of adults consists of these avoidance behaviors. If successive words in a sentence show increasing redundancy, it would be expected that the likelihood of stuttering would decline within a sentence.

One can gain support for this explanation from investigations which have shown that the general level of stuttering is some function of the social importance or meaningfulness of speech content. It has been shown, for example, that a group of subjects stuttered more frequently in reading aloud meaningful prose material than in reading aloud a series of numbers (Sheehan, 1958). A variety of other studies have reported similar findings (Brown & Moren, 1938; Eisenson & Horowitz, 1945). Apparently, then, the stutterer is responsive to the general social importance of his speech content in terms of a higher level of stuttering. It is not clear, however, whether this monitoring of social importance is carried out and acted upon at the word level, as Brown suggests.

In a recent study some evidence was obtained which indicated that Brown's view might not be an adequate explanation of the word-position gradient (Conway & Quarrington, 1963). The incidence of stuttering on initial, medial, and terminal words in seven-word sequences was studied under three levels of contextual organization. The extremes of contextual organization were normal English sentence structure and a random series of seven words. It was believed that these extremes would also represent extremes of a word-position redundancy gradient. The argument was made that, if Brown's views were correct, a gradient of stuttering incidence would be apparent in sentences characterized by a normal-redundancy gradient but relatively absent in lower levels of contextual organization. The results, however, indicated a comparable positional gradient of stuttering in all three levels of contextual organization studied. The general level of stuttering, nevertheless, did decline with lower levels of organization, which supports the notion that there is a general or situational responsiveness of stutterers to the social meaning of their speech efforts.

The present study extends the examination of the word-position gradient of stuttering incidence in the oral reading of a continuous prose passage. It was felt that earlier failure to demonstrate the operation of information value at the word level might be due to the restricted nature of the speech act studied (i.e., seven-word sequences), and that the relationship might be apparent in the extended speech situation which more closely resembled everyday speaking.

METHOD

A passage of banal prose, narrating a "walk by the river," was constructed, using commonplace syntactic structures and words occurring among the 1,050 most frequent words according to the Lorge-Thorndike count (Thorndike & Lorge, 1944). The complete piece of prose consisted of 95 words in nine sentences, the shortest being 6 words and the longest 14 words in length. The variable of word position was simply constructed by assigning each word a value corresponding to its numbered position in the sentence in which it occurred.

The information carried by each word was estimated by determining the extent to which it could be predicted by a population of normal adult speakers. This mode of measuring the information value of each word appeared appropriate in the light of Brown's discussion of "communicative importance." The method employed was fashioned after that developed by Selfridge (Miller, 1951). A device was constructed for successively exposing each word in the passage until the complete passage was revealed. The subjects were asked to guess the first word of the passage, and, having offered a guess, the word was shown to them. The subjects then proceeded to the second and subsequent words in this manner until guesses for each word had been obtained.

Forty-six fluent speakers served in this part of the investigation. The results showed a maximum range of expectedness with 3 words being correctly guessed by all subjects, while 17 words were not guessed by any subject. The mean percentage of correct guesses was 37.3. Higher scores in this variable imply, of course, lower information value.

TABLE 1

RELATIONS OF WORD POSITION AND PREDICTABILITY TO STUTTERING INCIDENCE

	Correlations with stuttering incidence		Regression weights
	First order	Partial	β
Word position (1)	−.49**	−.45**	−.441
Word predictability (2)	−.32**	−.25*	−.223

Note—$r(1)(2) = .22$*.
* $p \leq .05$.
** $p \leqq .01$.

A measure of stuttering incidence for each word was obtained by asking 24 adult stutterers, who were under investigation or recently admitted to treatment at the Toronto Psychiatric Hospital, to read aloud the experimental passage of prose. Each stutterer was instructed to read the passage as well as possible to the same auditor, and each performance was recorded on tape. Interjudge agreement between the experimenter and an experienced colleague was studied on judgments of performance in 10 subjects. The results showed high agreement with only one set of observations falling below the 95% agreement level.

Judged in this way, the range of stuttering incidence varied from 1 word stuttered by 16 subjects to 16 words on which no stuttering was observed. The mean incidence of stuttering was 3.57 persons, or 14.9% of the total group.

RESULTS

The three variables, word position, word predictability, and stuttering incidence were correlated over the 95 words of the experimental prose passage. The three first-order correlations are given in Table 1.

The product-moment coefficient of the correlation between word position and word predictability was .22, which is in line with the expectation that word predictability and redundancy should increase with successive word positions in a sentence. The relationship, however, is a rather weak one in this particular example.

The correlation between stuttering incidence and word predictability was −.32, while the correlation between stuttering incidence and word position was −.49. These correlation coefficients indicate that the incidence of stuttering increased with words of greater information value and with earlier sentence position.

A further examination of the relative importance of the two independent variables was carried out by determining their partial correlations with stuttering incidence. These correlations are shown in Table 1. The partial correlation of stuttering incidence and word predictability, with word position held constant, is −.25, a value differing significantly from zero at the .05 level of confidence. The corresponding coefficient of partial correlation for word position and stuttering incidence is $-.45$, which is significant at better than the .001 level of confidence. An additional glimpse of the relative strength of the dependencies may be found in the coefficients of the multiple regression of stuttering incidence on the two independent variables. These are given in their β form in Table 1.

It seems clear, then, that both word predictability and word position independently account for significant portions of the variance in stuttering incidence. The relationship of word position to stuttering incidence appears to be the more potent functional relationship.

DISCUSSION

The present findings confirm Brown's assertion that the likelihood that a word will be stuttered is some function of its information value in the total speech act. It was Brown's contention, however, that the positional gradient of stuttering in sentences could be explained by the general decline in the information value of successive words. This claim is not supported by the results obtained. It would appear that word position has a relationship to stuttering incidence independent of word information.

The present study was suggested by the interpretation of the word-position gradient, which issues from an approach-avoidance conflict view of stuttering (Sheehan, 1953, 1958). According to his fear-reduction hypothesis (Sheehan, 1958) the act of stuttering reduces the speech avoidance gradient and thus permits the eventual delivery of the feared word in a fluent fashion. It appears a simple extension of this argument to assume that the reduction of the avoidance gradient associated with stuttering tends to continue for a brief time so that the next word is approached with the point of maximum conflict nearer the point of word emission. In other words, the likelihood of stuttering on any particular word is some function of the amount of stuttering present on the preceding word or words. If stuttering has occurred on preceding words, then the likelihood of stuttering is decreased. Fluent delivery of words or silence, it is assumed, increases the avoidance gradient in the stutterer. This state of affairs accounts for the appearance of stuttering in cycles or waves, according to Sheehan.

The initial word of a sentence is approached with a high-avoidance gradient, which is in part a function of the preceding silence. The initial word is thus associated with a high likelihood of stuttering. If stuttering does in fact occur, the avoidance gradient is reduced as is the likelihood of stuttering on the following word. The positional gradient of stuttering in sentences is then due to the entrainment of a cycle of avoidance reduction in the typical act of sentence production. In this fashion stuttered words tend to occur at the beginning of a sentence, and fluent words toward the end, in a sentence of average length.

While one cycle of nonfluency may occur in a sentence of typical length, in sentences of greater length two or more cycles of nonfluency might occur. In a study of spontaneous speech (Hejna, 1955) some evidence has been found of a second phase of nonfluency beginning around words occupying the fifteenth and subsequent positions.

It should be observed that an alternative interpretation of the word-position gradient is possible in which this property of words is seen as yet another discriminative stimulus characteristic of words which may have a unique source and schedule of reinforcement.

The operant character of nonfluency in normal subjects, and dysfluency in stutterers has been demonstrated (Flanagan, Goldiamond & Azrin, 1958, 1959). An elaboration of this view has been offered which accounts for a good deal of what is known clinically with regard to the conditions which elicit stuttering (Shames & Sherrick, 1963).

In the matter of positional gradient further work is required in order to determine whether an adequate explanation resides in the positional property of words, or whether the positional phenomenon is incidental and is merely the reflection of a cycle of internal transitions of speech avoidance.

REFERENCES

BROWN, S. F. Stuttering with relation to word accent and word position. *Journal of Abnormal and Social Psychology*, 1938, 33, 112–120.

BROWN, S. F. An analysis of certain data concerning loci of "stutterings" from the viewpoint of general semantics. In M. Kendig (Ed.), *Papers from the Second American Congress on General Semantics*. Chicago: Institute of General Semantics, 1943.

BROWN, S. F. The loci of stutterings in the speech sequence. *Journal of Speech Disorders,* 1945, **10,** 181–192.

BROWN, S. F., & MOREN, A. A further study of stuttering in relation to various speech sounds. *Quarterly Journal of Speech,* 1938, **24,** 390–397.

CONWAY, J. K., & QUARRINGTON, B. J. Positional effects in the stuttering of contextually organized verbal material. *Journal of Abnormal and Social Psychology,* 1963, **67,** 299–303.

EISENSON, J., & HOROWITZ, ESTHER. The influence of propositionality on stuttering. *Journal of Speech Disorders,* 1945, **10,** 193–197.

FLANAGAN, B., GOLDIAMOND, I., & AZRIN, N. H. Operant stuttering: The control of stuttering behavior through response-contingent consequences. *Journal of the Experimental Analysis of Behavior,* 1958, **1,** 173–177.

FLANAGAN, B., GOLDIAMOND, I., & AZRIN, N. H. Instatement of stuttering in normally fluent individuals through operant procedures. *Science,* 1959, **130,** 979–981.

HEJNA, R. F. A study of the loci of stuttering in spontaneous speech. *Dissertation Abstracts,* 1955, **15,** 1674–1675.

MILLER, G. A. *Language and communication.* New York: McGraw-Hill, 1951.

SHAMES, G. H., & SHERRICK, C. E. A discussion of nonfluency and stuttering as operant behavior. *Journal of Speech Disorders,* 1963, **28,** 3–18.

SHEEHAN, J. G. Theory of stuttering as an approach-avoidance conflict. *Journal of Psychology,* 1953, **36,** 27–49.

SHEEHAN, J. G. Conflict theory of stuttering. In J. Eisenson (Ed.), *Stuttering: A symposium.* New York: Harper, 1958. Pp. 121–166.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received June 15, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 225–229

CASE REPORT

CLASSICAL AND OPERANT FACTORS IN THE TREATMENT OF A SCHOOL PHOBIA[1]

ARNOLD A. LAZARUS,[2] GERALD C. DAVISON, AND DAVID A. POLEFKA
Stanford University

To the best of our knowledge, this report is the 1st to recognize the advantages of employing both classical and operant conditioning procedures in the treatment of a neurotic case. A model which appeared to have heuristic value was developed: When avoidance behavior is motivated by high levels of anxiety, classical counterconditioning techniques are called for; when anxiety is minimal, and avoidance behavior is seemingly maintained by various secondary reinforcers, operant strategies should be applied. Furthermore, this paper indicates that the practice of interchanging therapists not only failed to disrupt or impede therapeutic progress but had certain distinct advantages.

Although the formal application of "learning theory" to clinical problems is widespread, the literature on this topic reflects a basic cleavage. Wolpe (1958) and Eysenck (1960) typify the use of the classical conditioning paradigm in the treatment of neurotic disorders, while Lindsley and Skinner (1954), King, Merrell, Lovinger, and Denny (1957), and Ferster (1961) exemplify the use of operant conditioning in the treatment of psychotic behavior. On the assumption that both "operants" and "respondents" enter into all therapeutic processes, the writers hypothesized that the deliberate and strategic use of both classical and operant conditioning procedures would have greater therapeutic effect than exclusive reliance on techniques derived from either procedure alone. The therapeutic utility of this rationale became obvious in the treatment of a severely disturbed (nonpsychotic) school-phobic child.

Strategy in "behavior therapy" consists essentially of introducing reinforcement contingencies that encourage the emergence of nondeviant response patterns. This may be achieved by pairing the reinforcer with a *stimulus* (as is the case in classical conditioning) and/or by making the reinforcer contingent upon a *response* (as is the case in operant conditioning). Apart from Patterson's (1965) successful application of predominantly operant techniques to a school-phobic child, the treatment of children's phobias by conditioning methods has hitherto relied almost exclusively on the classical paradigm (Bentler, 1962; Jones, 1924; Lazarus, 1960; Lazarus & Abramovitz, 1962; Lazarus & Rachman, 1957; Wolpe, 1958). It could be argued, however, that some of the above-named investigators made inadvertent use of the operant rubric. In a case alluded to by Lazarus and Abramovitz (1962) for instance, a child with "widespread areas of disturbance" failed to benefit from counterconditioning therapy but required "broader therapeutic handling." The therapeutic mainstay in this instance actually amounted to persuading the parents to alter certain of their actions which were sustaining their child's deviant responses (i.e., an operant strategy). The reapplication of counterconditioning techniques then effected a rapid recovery. The present paper is an endeavor to illustrate how the deliberate (rather than inadvertent) use of these two theoretical models at crucial phases throughout treatment proved therapeutically expeditious.

CASE STUDY

History of the Problem

When he was referred for therapy Paul, age 9, had been absent from school for 3 weeks. The summer vacation had ended 6 weeks previously, and on entering the fourth grade, Paul avoided the classroom situation. He was often found hiding in the cloakroom, and subsequently began spending less time at school each day. Thereafter, neither threats, bribes, nor punishments could induce him to re-enter school.

Paul's history revealed a series of similar episodes. During his first day of kindergarten he succeeded in climbing over an extremely high wall and fled home. His first-grade teacher considered him to be "disturbed." Serious difficulties regarding school attendance were first exhibited when Paul entered the second grade of a parochial school. It was alleged that the second-grade teacher who, according to Paul,

---

[1] Reprints may be obtained by writing to the second author at Stanford University.

[2] Now at the Witwatersrand University Medical School, Johannesburg, South Africa.

"looked like a witch," generally intimidated the children and was very free with physical punishment.[3] Paul retrospectively informed his parents that he felt as though "the devil was in the classroom." At this stage he became progressively more reluctant to enter the school and finally refused entirely. A psychiatrist was consulted and is reported to have advised the parents to use coercion, whereupon Paul was literally dragged screaming to school by a truant officer. Paul was especially bitter about his experience with the psychiatrist. "All we did was talk and then the truant officer came." In the third grade Paul was transferred to the neighborhood public school where he spent a trouble-free year at the hands of an exceedingly kind teacher.

Family History

Paul was the fourth of eight children, the first boy in a devout, orthodox Roman Catholic family. His sisters were aged 14, 13, 11, 7, and 6 years, respectively; his two brothers were 8 and 2½ years old. The father was a moody, anxiously ambitious electronics engineer who had insight into the fact that his subjective occupational insecurities intruded into the home. A harsh disciplinarian—"I run a tight ship"—he impulsively meted out punishment for any act which deviated even slightly from his perfectionistic standards. He found it significant that Paul, of all the children, was particularly sensitive to his moods, and described himself as being "especially close to Paul" while commenting that "he rarely tells me things." In his desire to protect his family from everyday hazards he was inclined to emphasize extreme consequences: "Don't touch that fluorescent bulb, son; there's poison in it and it will kill you!"

The mother, although openly affectionate and less rigid and demanding than her husband, took pains to respond towards her eight children in an unbiased fashion. She stressed, however, that "Paul touches my nerve center," and stated that they frequently quarreled in the father's absence. She had always found Paul "less cuddlesome" than his siblings. When he was 2 years old, she would lock him out of the house, "so as to develop his independence." It is significant to note that this occurred immediately following the birth of his first brother. In general, she was inclined to be inconsistent when administering rewards and punishment. Psychometric testing suggested that Paul was uncertain whether a given response would meet with criticism and rejection or kind attention from his mother. It was nevertheless evident that Paul was eager to receive a greater share of his mother's highly-rationed time.

The lad himself was somewhat small and frail-looking. Although reticent, essentially aloof and somewhat withdrawn, he was capable of unexpected vigor and self-assertion when he chose to participate in sporting activities. From the outset, the therapists noted his labile and expressive reactions to all stressful stimuli. The extent of his subjective discomfort was easily gauged by clearly discernible responses. As the magnitude of anxiety increased, there was a concomitant progression of overt signs—increased reticence, a postural stoop, a general constriction of movement, tearfilled eyes, mild trembling, pronounced blanching, culminating in sobbing and immobility. As will be shown below, these emotional indices were crucial in selecting appropriate therapeutic strategies.

A series of specific traumatic events commenced with his near-drowning when 5 years old. Towards the end of his third grade, he underwent a serious appendectomy with critical complications, which was followed by painful postoperative experiences in a doctor's consulting room. During one of these examinations, as Paul bitterly recounted, he had been left alone by his parents. Shortly after his recovery from surgery, he witnessed a drowning which upset him considerably. Following his entry into the fourth grade, the sudden death of a 12-year-old girl, who had been a close friend of his elder sister, profoundly affected the entire family. It is also noteworthy that Paul's father experienced personal stress in his work situation during the child's turbulent second grade, as well as immediately preceding fourth grade. Finally, Paul seemed to have been intimidated by a warning from his eldest sister that fourth grade school work was particularly difficult.

Therapeutic Procedure

After the initial interview, it was evident that Paul's school phobia was the most disruptive response pattern of a generally bewildered and intimidated child. Although subsequent interviews revealed the plethora of familial tensions, situational crises, and specific traumatic events outlined above, the initial therapeutic objective was to reinstate normal school attendance. Nevertheless it was clearly apparent that the home situation in general, and more particularly, specific examples of parental mishandling would ultimately require therapeutic intervention.

The application of numerous techniques in the consulting room (e.g., systematic desensitization[4]) was abandoned because of the child's inarticulateness and acquiescent response tendency. It was obvious that his verbal reports were aimed at eliciting approval rather than describing his true feelings. Desensitization *in vivo* was therefore employed as the principal therapeutic strategy.

The school was situated two and one half blocks away from the home. The routine was for Paul to leave for school at 8:30 A.M. in order to arrive by 8:40. The first recess was from 10:00–10:30; lunch break from 12:00–1:00; and classes ended at 3:30 P.M. At the time when therapy was initiated, the boy was extremely surly and dejected in the mornings (as reported by the parents), refused breakfast, rarely dressed himself, and became noticeably more fearful toward 8:30. Parental attempts at reassurance, coaxing, or coercion elicited only sobbing and further withdrawal.

---

[3] Vehement complaints from many parents finally led to the dismissal of this teacher.

[4] Systematic desensitization entails the presentation of carefully graded situations, which are subjectively noxious, to the imagination of a deeply relaxed patient until the most personally distressing events no longer evoke any anxiety (see Wolpe, 1961).

Accordingly, the boy was exposed to the following increasingly difficult steps along the main dimensions of his school phobia:

1. On a Sunday afternoon, accompanied by the therapists, he walked from his house to the school. The therapists were able to allay Paul's anxiety by means of distraction and humor, so that his initial exposure was relatively pleasant.

2. On the next 2 days at 8:30 A.M., accompanied by one of the therapists, he walked from his house into the schoolyard. Again, Paul's feelings of anxiety were reduced by means of coaxing, encouragement, relaxation, and the use of "emotive imagery" (i.e., the deliberate picturing of subjectively pleasant images such as Christmas and a visit to Disneyland, while relating them to the school situation; see Lazarus & Abramovitz, 1962). Approximately 15 minutes were spent roaming around the school grounds, after which Paul returned home.

3. After school was over for the day, the therapist was able to persuade the boy to enter the classroom and sit down at his desk. Part of the normal school routine was then playfully enacted.

4. On the following three mornings, the therapist accompanied the boy into the classroom with the other children. They chatted with the teacher, and left immediately after the opening exercises.

5. A week after beginning this program, Paul spent the entire morning in class. The therapist sat in the classroom and smiled approvingly at Paul whenever he interacted with his classmates or the teacher. After eating his lunch he participated in an active ball game, and returned to his house with the therapist at 12:30. (Since parent-teacher conferences were held during that entire week, afternoon classes were discontinued.)

6. Two days later when Paul and the therapist arrived at school, the boy lined up with the other children and allowed the therapist to wait for him inside the classroom. This was the first time that Paul had not insisted on having the therapist in constant view.

7. Thereafter, the therapist sat in the school library adjoining the classroom.

8. It was then agreed that the therapist would leave at 2:30 P.M. while Paul remained for the last hour of school.

9. On the following day, Paul remained alone at school from 1:45 P.M. until 2:45 P.M. (Earlier that day, the therapist had unsuccessfully attempted to leave the boy alone from 10 until noon.)

10. Instead of fetching the boy at his home, the therapist arranged to meet him at the school gate at 8:30 A.M. Paul also agreed to remain alone at school from 10:45 A.M. until noon provided that the therapist return to eat lunch with him. At 1:45 P.M. the therapist left again with the promise that if the boy remained until school ended (3:30 P.M.) he would visit Paul that evening and play the guitar for him.

11. Occasional setbacks made it necessary to instruct the lad's mother not to allow the boy into the house during school hours. In addition, the teacher was asked to provide special jobs for the boy so as to increase his active participation and make school more attractive.

12. The family doctor was asked to prescribe a mild tranquilizer for the boy to take on awakening so as to reduce his anticipatory anxieties.

13. After meeting the boy in the mornings, the therapist gradually left him alone at school for progressively longer periods of time. After 6 days of this procedure, the therapist was able to leave at 10 A.M.

14. The boy was assured that the therapist would be in the faculty room until 10 A.M., if needed. Thus, he came to school knowing the therapist was present, but not actually seeing him.

15. With Paul's consent the therapist arrived at school shortly *after* the boy entered the classroom at 8:40 A.M.

16. School attendance independent of the therapist's presence was achieved by means of specific rewards (a comic book, and variously colored tokens which would eventually procure a baseball glove) contingent upon his entering school and remaining there alone. He was at liberty to telephone the therapist in the morning if he wanted him at school, in which event he would forfeit his rewards for that day.

17. Since the therapist's presence seemed to have at least as much reward value as the comic books and tokens, it was necessary to enlist the mother's cooperation to effect the therapist's final withdrawal. The overall diminution of the boy's anxieties, together with general gains which had accrued to his home situation, made it therapeutically feasible for the mother to emphasize the fact that school attendance was compulsory, and that social agencies beyond the control of both therapists and parents would enforce this requirement eventually.

18. Approximately 3 weeks later, Paul had accumulated enough tokens to procure his baseball glove. He then agreed with his parents that rewards of this kind were no longer necessary.

THEORETICAL IMPLICATIONS

It should not be inferred that Paul's improvement followed a smooth monotonic progression. Numerous setbacks of varying degrees of severity occurred throughout the entire treatment program, which extended over 4½ months. These episodes were differentially handled depending upon the therapist's assessment of the child's anxiety at that time, and his judgment of the degree to which the boy had mastered the preceding therapeutic steps.

It became apparent that the school phobia was comprised of two separate factors: (*a*) avoidance behavior motivated by intense fear of the school situation, and (*b*) avoidance behavior maintained by various secondary reinforcers, mainly attention from parents, siblings, and therapists. During the initial phases of therapy, the boy's high level of anxiety dictated the use of reciprocal inhibition methods (Wolpe, 1958). The therapists actively inhibited the boy's anxiety elicited by various aspects of the school setting (as in Step 2 above). The later stages of therapy were characterized by a decrease in Paul's overall anxiety without a concomitant decrease in avoidance behavior. After Step 15 the boy appeared to be minimally anxious. An operant strategy which made various rewards contingent on school attendance was therefore selected.

A Proposed Model

Although the division between classical and operant procedures became clearly discernible towards the terminal phases of treatment, many situations arose which necessitated the deliberate choice of one or other paradigm. A model was developed for determining when each was likely to prove maximally effective. On several occasions for instance, Paul left the classroom, entered the library and told the therapist, "I'm scared." At this point the choice of strategy became crucial. In strict operant terms, active attempts to reduce anxiety by means of attention and reassurance would reinforce classroom-leaving behavior. On the other hand, the classical paradigm would predict that to withhold immediate attention and make it contingent upon returning to the classroom would augment the child's anxiety and thus reinforce avoidance behavior. The critical factor in determining the appropriate procedure was the degree of anxiety as judged by the therapist.

An inappropriate use of the operant model could prove anti-therapeutic. If the level of anxiety is very high, a premature re-exposure to the feared situation will probably lead to increased sensitivity. Moreover, if this heightened level of anxiety leads to another escape response, the resultant anxiety-reduction will strengthen the avoidance responses (classroom-leaving behavior in this instance). It was also reasoned that when highly anxious, the boy would be unable to attend to the teacher, interact with his peers, or make any other responses which ordinarily reduced his anxiety.

An inappropriate use of the classical model would also impede therapeutic progress. The very acts of inducing relaxation, employing "emotive imagery," and giving reassurance may provide positive reinforcement for dependent behavior. The afore-mentioned difficulties in 'phasing out' may be attributed to this possible side effect of in vivo desensitization. The gains which accrue when high levels of anxiety are thus decreased, however, temporarily out-weigh the disadvantages of increased dependency.

DISCUSSION

It may be argued that a disproportionate amount of time and effort was expended in attaining the principal therapeutic objective, viz., normal school attendance. Urgent cases, however, who are neurotically incapacitated and unamenable to interview techniques, would seem to require therapeutic intervention beyond the confines of the consulting room. It should be emphasized that school phobia in a child is almost as pressing and disruptive a problem as occupational fears in an adult.

Since therapy *in vivo* makes heavy demands on the therapist's time, the senior author decided to enlist the assistance of two graduate students in clinical psychology (the co-authors). During the first exposure to school (see Step 1) Paul rapidly developed an attachment to one of the cotherapists (G.C.D.). As the application of reciprocal inhibition methods is conceivably facilitated by the nonspecific anxiety-inhibiting effects of a "good relationship" (Lazarus, 1961), this therapist carried out the first eight steps. Thereafter, the choice of therapist was partly determined by the academic and clinical commitments of the respective authors. Significantly, very little disturbance was occasioned by the constant change of therapists. There was often a distinct advantage in being able to alternate therapists; it was found, following a setback, that it was helpful to change therapists in order to offset the negative effects of being associated with sensitizing experiences.

The adjunctive use of a tranquilizer seemed to be of limited therapeutic value. Initially, it appeared to reduce the boy's anticipatory anxieties, but the absence of negative effects whenever it was forgotten, suggests that a placebo would have been as effective.

The therapists kept in close communication with the parents, who were encouraged to telephone whenever situational crises arose. As soon as normal school attendance had been more or less reinstated (Step 14), the therapists held a "family conference." In the main, the implications of the father's harsh and restrictive tendencies, along with the mother's inconsistent and ambivalent attitudes, were made clear to them. A long list of specific "do's" and "dont's" was drawn up and discussed. Apart from minor points of disagreement, the parents responded in an intelligent and receptive manner and subsequently implemented many of the recommendations.

According to the mother's reports, Paul's behavior also improved in areas outside of the school situation. She referred to his marked decrease in moodiness, his increased willingness to participate in household chores, more congenial relationships with his peers, and general gains in self-sufficiency.

Ten months after the termination of therapy, a follow-up inquiry revealed that Paul had not only maintained his gains, but had made further progress.

REFERENCES

BENTLER, P. M. An infant's phobia treated with reciprocal inhibition therapy. *Journal of Child Psychology and Psychiatry*, 1962, **3**, 185–189.

EYSENCK, H. J. (Ed.) *Behaviour therapy and the neuroses*. New York: Pergamon Press, 1960.

FERSTER, C. B. Positive reinforcement and behavioral deficits in autistic children. *Child Development*, 1961, **32**, 437–456.

JONES, MARY C. Elimination of children's fears. *Journal of Experimental Psychology*, 1924, **7**, 383–390.

KING, G. F., MERRELL, D., LOVINGER, E., & DENNY, M. Operant motor behavior in acute schizophrenics. *Journal of Personality*, 1957, **25**, 317–326.

LAZARUS, A. A. The elimination of children's phobias by deconditioning. In H. J. Eysenck (Ed.), *Behaviour therapy and the neuroses*. New York: Pergamon Press, 1960.

LAZARUS, A. A. Group therapy of phobic disorders by systematic desensitization. *Journal of Abnormal and Social Psychology*, 1961, **63**, 504–510.

LAZARUS, A. A., & ABRAMOVITZ, A. The use of "emotive imagery" in the treatment of children's phobias. *Journal of Mental Science*, 1962, **108**, 191–195.

LAZARUS, A. A., & RACHMAN, S. The use of systematic desensitization in psychotherapy. *South African Medical Journal*, 1957, **31**, 334–337.

LINDSLEY, O. R., & SKINNER, B. F. A method for the experimental analysis of the behavior of psychotic patients. *American Psychologist*, 1954, **9**, 419–420.

PATTERSON, G. R. A learning theory approach to the treatment of the school phobic child. In L. Ullman & L. Krasner (Eds.), *Case studies in behavior modification*. New York: Holt, Rinehart, & Winston, 1965.

WOLPE, J. *Psychotherapy by reciprocal inhibition.* Stanford: Stanford Univer. Press, 1958.

WOLPE, J. The systematic desensitization treatment of neuroses. *Journal of Nervous and Mental Disease*, 1961, **132**, 189–203.

(Received May 25, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 3, 230

FAMILY SIZE IN ALCOHOLISM

COLIN M. SMITH
University Hospital, Saskatoon, Canada

Sprott (1964) has recently taken issue with Smart's (1963) use of chi-square in relation to his data on the size of the family of origin among alcoholics. Smart (1964) reviewing Sprott's paper states that the "suggested changes in the family size analyses do not affect the conclusion that persons from large families are over-represented among alcoholics."

Sprott's criticisms are well founded but both authors apparently fail to notice that the above conclusion is completely unwarranted by the data. Smart (1963) gives the mean age of the alcoholics as 43.02 years. Presumably all or very nearly all the patients are from *completed* families of origin; indeed, otherwise the analysis for birth order used would be quite misleading. The expected values for family size are stated to be taken from the 1931 census for Ontario (Census for Canada 1931, Table 96, p. 1314). A family for the purposes of this census consists of husband and wife, or parents with sons or daughters living together at the time of the survey (Census, pp. xxi, 1174, 1290). Such a family is not of course necessarily either completed or even intact. Thus a young couple with one child will be recorded as such; however, over the next 15 or 20 years they may add four or five more children to their family. On the other hand, a middle-aged couple three of whose five children have left home will be listed as forming a family of two children. It must be obvious that a comparison of this kind is completely invalid and could only lead to an apparent overrepresentation of larger families of origin among the alcoholic group and this in fact is what was found.

It is of interest to compare Smart's data to that reported by Smith & McIntyre (1963) in 106 Saskatchewan patients admitted to a medical ward of a general hospital. The average age (37.5 years) of the medical controls differs slightly from that of Smart's alcoholics. While it cannot be assumed that the controls are representative of the population from which Smart's patients were drawn it is worthy of note that the differences between the two groups are actually in the *opposite* direction to that predicted by Smart's hypothesis, the alcoholics showing a smaller proportion of larger sibships (see Table 1).

TABLE 1

SIBSHIPS OF 242 ALCOHOLICS COMPARED WITH THOSE OF 106 MEDICAL CONTROLS

Size of sibship	Alcoholics (Smart's data)	Medical controls (Smith's data)
1	21	4
2	32	10
3	40	13
4	47	11
5 or more	102	68
	242	106

Note.—$\chi^2 = 15.12$, $df = 4$, $p < .005$.

REFERENCES

DOMINION BUREAU OF STATISTICS. *The seventh census of Canada*. Vol. 5, Ottawa: D.B.S., 1931.

SMART, R. G. Alcoholism, birth order and family size. *Journal of Abnormal and Social Psychology*, 1963, **66**, 17–23.

SMART, R. G. A response to Sprotts "Use of Chi Square." *Journal of Abnormal and Social Psychology*, 1964, **69**, 103–105.

SMITH, C. M., & MCINTYRE, S. Family size, birth rank and ordinal position in psychiatric illness. *Canadian Psychiatric Association Journal*, 1963, **8**, 244–248.

SPROTT, D. A. Use of chi square. *Journal of Abnormal and Social Psychology*, 1964, **69**, 101–103.

(Received October 23, 1964)

Academic Press

Deranged Memory

A Psychonomic Study of the Amnesic Syndrome

by *George A. Talland*

Deranged Memory is an account of the most extensive investigation ever conducted with a group of patients in the chronic state of Korsakoff's Disease. The study, composed of an examination of more than twenty patients on forty or more occasions, was designed to test specific hypotheses about disturbances characteristic of the amnesic syndrome and, indirectly, the processes involved in normal memory function. Hypotheses related to certain theories of psychology and neurophysiology were subject to critical study.

The book includes:

- a review of the literature published in five languages on the psychopathology of the Korsakoff syndrome
- data on the performance capacity of the patients in over 100 experimental tasks
- critical examinations of confabulation in both chronic and acute phases of the amnesic syndrome
- an empirical evaluation of certain explanatory models proposed by other investigators and by the author himself

CONTENTS: INTRODUCTORY: The Goals and Methods of the Research. CLINICAL: The Amnesic Patient. Confabulation. HISTORICAL: The Literature of the Amnesic Syndrome. EXPERIMENTAL: Verbal and Motor Skills: Activity Rate. Perception. Intelligence; Reasoning. Learning and Remembering. Reproduction and Distortion of Narrative Texts. THEORETICAL: Memory Disturbance. Temporal Signs and Self-Reference. Premature Closure of Activation. *References. Author Index—Subject Index.*

(*T086*) May 1965, 356 pp., $9.50

Progress in Experimental Personality Research

VOLUME 1

edited by *Brendan A. Maher*

This new serial publication is designed for those who require knowledge of recent developments in the study of personality, particularly psychologists, psychopathologists, psychiatrists, sociologists, educators, and students of behavioral sciences. Contributions range from original works in important technical areas to syntheses of data and to integrated summaries of present knowledge.

In this first volume, papers deal with interpersonal perception, experimenter-bias, the personality correlates of repression and sensitization, conditioning and personality, and processes in projective perception. Two original and systematic contributions in the understanding of schizophrenia are of particular significance.

CONTENTS: P. H. VENABLES, Input Dysfunction in Schizophrenia. LOREN J. CHAPMAN, JEAN P. CHAPMAN, and GLENN A. MILLER, A Theory of Verbal Behavior in Schizophrenia. ROBERT ROSENTHAL, The Effect of the Experimenter on the Results of Psychological Research. S. H. LOVIBOND, Personality and Conditioning. DONN BYRNE, Repression-Sensitization as a Dimension of Personality. VICTOR B. CLINE, Interpersonal Perception. DOUGLAS T. KENNY, Stimulus Functions in Projective Techniques. *References. Author Index—Subject Index.*

(*P750*) 1964, 368 pp., $9.50

*** A Continuation Order** . . . *assures you of having the latest information in your subject as soon as it is available. We recommend the placing of a continuation order for serial publications, series, and journals as it will authorize us to ship and bill automatically upon publication. The order may be cancelled at any time.*

Academic Press NEW YORK AND LONDON • 111 FIFTH AVENUE NEW YORK N Y 10003 • BERKELEY SQUARE HOUSE LONDON W 1 • PHONE 212-OR 7-6713

MATHEMATICS • PHYSICS • CHEMISTRY • BIOLOGICAL SCIENCES • MEDICAL RESEARCH • SPACE SCIENCES • ENGINEERING • PSYCHOLOGY

The Journal of conflict RESOLUTION

A quarterly for research related to war and peace

JUNE 1965 ISSUE (IX, 2):

Military capabilities and bargaining behavior . . . . *Lloyd Jensen*

Aging, time perspective, and preferred solutions to international conflicts . . . . *Kenneth Gergen and Kurt Back*

On the morality of chemical/biological war . . . . *Richard Krickus*

Other articles; gaming section; book reviews

$5.50 per year, $2.00 per issue

Published by The Center for Research on Conflict Resolution
The University of Michigan, Ann Arbor, Michigan 48104

We call your attention to the following recent Special Issues of the AMERICAN ANTHROPOLOGIST

Selected Papers in Method and Technique, Vol. 65, No. 5, October, 1963. Proxemic behavior, ethnographic and mathematical models, computer programming, ranking method, and formal ethnographic analysis.

Transcultural Studies in Cognition, Vol. 66, No. 3, Part 2, June, 1964. Ethnolinguistic theory, semantics, ethnoscience, cognition, and method in cross-cultural studies.

New Guinea: The Central Highlands, Vol. 66, No. 4, Part 2, August, 1964. Original papers on ecology, prehistory, language, social system, warfare, male-female relationships, administration, politics, religion, economy, and acculturation.

The Ethnography of Communication, Vol. 66, No. 6, Part 2, December, 1964. Social origins of codes, proverbs, language and listener, baby talk, words for self, linguistic and social interaction, correlates of social stratification, speech communities.

$2.75 each

Send remittance to:

AMERICAN ANTHROPOLOGICAL ASSOCIATION
1530 P Street, N. W.
Washington, D. C. 20005

Volume 70 August 1965 Number 4

Journal of Abnormal Psychology

Editor: Howard F. Hunt, *New York State Psychiatric Institute and Columbia University*

Associate Editor: William N. Thetford, *College of Physicians and Surgeons, Columbia University*

Consulting Editors

Theodore X. Barber
Medfield Research Foundation

Harold Basowitz
New York University

Eugene I. Burdock
New York State Psychiatric Institute and Columbia University

John M. Butler
University of Chicago

Arthur C. Carr
New York State Psychiatric Institute and Columbia University

Loren J. Chapman
Southern Illinois University

John H. Cumming
New York State Department of Mental Hygiene

W. Grant Dahlstrom
University of North Carolina

Jerome D. Frank
Johns Hopkins University

Sol L. Garfield
Columbia University

Ann Magaret Garner
Lincoln, Nebraska

Israel Goldiamond
Institute for Behavioral Research

Harrison G. Gough
University of California, Berkeley

Starke R. Hathaway
University of Minnesota

Ralph Heine
University of Chicago

Nicholas Hobbs
George Peabody College

Lawrence C. Kolb
New York State Psychiatric Institute and Columbia University

Sheldon J. Korchin
University of California

William G. Lawlor
Fordham University

Ardie Lubin
U. S. Naval Hospital, San Diego

Salvatore Maddi
University of Chicago

David Rosenthal
National Institute of Mental Health

Alan K. Rosenwald
Illinois State Psychiatric Institute

Saul Rosenzweig
Washington University, St. Louis

Jacques Rutschmann
New York State Psychiatric Institute

Seymour Sarason
Yale University

Joseph Schachter
New York State Psychiatric Institute and Columbia University

Roy Schafer
Yale University

John Shlien
University of Chicago

Morris I. Stein
New York University

Charles Strother
University of Washington

Harold L. Williams
University of Oklahoma School of Medicine

C. L. Winder
Michigan State University

Joseph Zubin
New York State Psychiatric Institute and Columbia University

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

FRANCES L. BREWER
Editorial Assistant

ELIZABETH S. REED
Advertising Manager

ROBERT L. HOOD
Circulation Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 231–243

MENTAL ACTIVITY AT SLEEP ONSET[1]

DAVID FOULKES AND GERALD VOGEL
University of Wyoming *University of Chicago*

Experimental data, based upon a total of 212 hypnagogic awakenings contributed by 9 Ss, are presented on the incidence of reports of various forms of hypnagogic mentation and on the relationship of such reports to EEG activity. It was found that mental activity in some form or another was invariably reported throughout the hypnagogic period. Content reported was predominantly visual and lacking in affective intensity and it became increasingly hallucinatory and unamenable to voluntary control as EEG patterns shifted from an alpha EEG to a spindling EEG. Dreamlike reports very similar to REM reports (i.e., dramatic, hallucinatory episodes) were quite common during descending EEG Stages 1 and 2, and, surprisingly enough, also occurred with an alpha rhythm.

Research by Aserinsky, Kleitman, and Dement (Aserinsky & Kleitman, 1955; Dement & Kleitman, 1957a, 1957b) has established that nocturnal dreaming occurs primarily during cyclically recurring periods of sleep characterized by a random, low-voltage EEG (EEG Stage 1) and by intermittent conjugate rapid eye movements (REMs) which are apparently related to the direction of dream imagery (Roffwarg, Dement, Muzio, & Fisher, 1962). Although mental content may be reported from other stages of sleep, that is, from non-REM EEG Stages 2, 3, and 4 in which 14 cycles per second sleep spindles and 1–3 cycles per second delta waves predominate, such content is generally more thoughtlike than dreamlike (Foulkes, 1962; Rechtschaffen, Verdone, & Wheaton, 1963) and may readily be discriminated from the typical Stage 1 REM-period dream (Monroe, Rechtschaffen, Foulkes, & Jensen, in press).

According to the electroencephalograph, the first cycle of a normal night's sleep consists of the progression from relaxed wakefulness, in which a 10 cycles per second alpha rhythm dominates the EEG record, quickly through Stages 1 and 2 to a period of Stages 3 and/or 4 which may last for an hour or more, and then a reasonably rapid return through Stage 2 to a period of "emergent" Stage 1, accompanied by REMs. The first REM period of the night typically occurs 90–100 minutes after sleep onset, and is of approximately 10-minutes duration. At its conclusion, a second cycle of EEG Stages 2 to 3–4 to 2 to 1 then occurs, with the Stage 1 REM period terminating this cycle averaging approximately 20 minutes in length. A 7- or 8-hour sleep period generally includes four or so of these cycles, with REM periods constituting a progressively larger proportion of each successive cycle.

Sleep-onset, or "descending," Stage 1 has been found to differ in a number of important respects from the emergent, or "ascending," Stage 1 which terminates each sleep cycle. It is unaccompanied by REMs, may instead be characterized by intermittent slow eye movements (SEMs) which lack any fixational quality and are sometimes binocularly asymmetrical (Aserinsky & Kleitman, 1955). Dement and Kleitman (1957a) awakened subjects from descending Stage 1 and found that disorganized flashes of visual imagery were reported, rather than dreams. Sleep-onset Stage 1 also has considerably lower auditory arousal thresholds than emergent Stage 1 (Dement & Kleitman, 1957a), lacks the saw-toothed EEG waves which are seen during REM periods, and demonstrates greater cortical reactivity to external stimuli in the form of evoked potentials than does the emergent Stage 1 REM period (Williams, Tepas, & Morlock, 1962). Dement (in press) has concluded that the emergent Stage 1 REM period is a unique phase of sleep, qualitatively different in its physiological manifestations and its experiential correlates both from non-

---

[1] This investigation was supported by grant M-04151 from the National Institute of Mental Health and by P. W. Project Number 17-165 from the State of Illinois.

REM sleep (EEG Stages 2, 3, and 4) and from sleep-onset Stage 1. He asserts that the emergent Stage 1 REM period is the organismic dreaming state, and that the amount of dreaming experienced by the subject may be accurately gauged from the number of uninterrupted REM episodes occurring through the night.

The identification of dreaming with periods of emergent EEG Stage 1 accompanied by REMs and the insistence that this "unique" phase of sleep is alone capable of sustaining dreaming implies that dreaming is absent at sleep onset. Maron, Rechtschaffen, and Wolpert (1964) failed to note REM periods occurring at the onset of daytime naps and it is a common observation that, except in the case of narcoleptics (Vogel, 1960; Rechtschaffen, Wolpert, Dement, Mitchell, & Fisher, 1963), REM periods never occur at the onset of night-time sleep. And yet, Maron et al. note that it is by no means infrequent for subjects to assert that they have dreamed as they fall off to sleep during daytime napping, and there are clear indications in the non-EEG literature on mental activity during sleep (e.g., Archer, 1947; Maury, 1878; Slight, 1924) that dreaming *does* occur at sleep onset. In the present study we have employed Dement and Kleitman's EEG and eye-movement recording technique in an examination of mental activity at sleep onset to try to resolve this discrepancy between contemporary psychophysiological theory, which has stressed the REM period as the locus of dreaming, and everyday and anecdotal observation, which has held that dreams may occur at sleep onset.

PROCEDURE

The subjects were 9 young adults, 8 males and 1 female. All were college students or college-educated and, as assessed by MMPIs and our own observations, free from any signs of gross psychological disturbance. Each subject was run for four experimental sessions, always conducted in the evening close to the subject's normal bedtime. Each session lasted for 2½–4 hours.

Upon his arrival at the laboratory, the subject put on his sleeping clothes. The experimenter then attached EEG and EOG (electrooculogram) electrodes to the subject's face and scalp. Standard (Dement & Kleitman, 1957a, 1957b) EEG electrode placements and procedures were followed, and EOG recordings were taken in monopolar fashion from electrodes at the outer canthus of each eye.

The subjects went to bed in a darkened, reasonably soundproof room. When the polygraph recording, monitored by the experimenter in another room, indicated an appropriate EEG and EOG pattern, the subject was called by name over an intercom system, and, upon some response, was asked, "What were you experiencing just before I called you?" The subjects had been instructed, prior to retiring, that the experimenter's call did not imply that the subject was supposed to be asleep or to have any experience to report, but rather that these were the topics to be investigated in the study. The subjects had been further instructed that if they felt they were experiencing nothing in particular prior to the experimenter's call, they should report this fact, since that would be a finding of some importance in its own right. The opening question was phrased in the broadest terms to try to prevent the subject from suppressing any content because it did not seem to him to be pertinent to the experimenter's interest, and was meant to elicit an initial spontaneous report from the subject.

Upon completion of his spontaneous report, the subject was asked a number of standardized questions concerning his state of wakefulness or sleep and the sensory imagery, hallucinatory quality, etc., of any material he may have reported. The questions included the following:

> What was taking place *just* at the moment when I called?
>
> At that moment would you say that you were: awake and reasonably alert; awake but drowsy; drifting off to sleep; in light sleep; in deep sleep?
>
> Was there any visual imagery? Describe it.
>
> [If only one image reported] Did the image seem to change or develop from the first time you saw it?
>
> What thoughts or feelings were going through your mind during the experience you have reported?
>
> In this experience did you hear, taste, or smell anything?
>
> In this experience did you perceive any bodily sensations (such as touch, movement, position, pressure, pain, bodily pleasure)?
>
> Did you experience any distortions in the way you sensed your own body or its parts (such as changes in size, shape, weight, or relationship of the parts)?
>
> Were there any distortions in the way familiar people or objects were represented?
>
> Were you an active participant in what you experienced, or just passively observing it?
>
> Were there any other persons in this experience?
>
> Were there any similarities between this experience and an ordinary daydream, nightdream, memory, or waking thought that you might have?
>
> Did you feel that you were controlling the course of your thoughts/imagery in the sense that you felt you were conjuring them up and could stop them if you wanted to?

> Were you, while the experience was occurring, aware that you were here in the laboratory, lying in your bed?
>
> Were you, during the experience, aware that you were observing the contents of your own mind, or did you feel that you were observing or participating in events out in the real world?

Following completion of this interview, the subject was allowed to attempt to return to sleep, and, when an appropriate EEG and EOG pattern was again achieved, the subject was again "awakened" and asked what he had been experiencing immediately prior to the experimenter's call. At the completion of the sixth interview, the experimental session was terminated.

Experimental sessions were generally scheduled 1 week apart, never on successive evenings. Except in the case of one subject, the same experimenter (one or the other of the authors or William Zimmerman, whose assistance is gratefully acknowledged) was assigned to all four of the sessions for any one subject.

Four categories of "awakening" were employed, reflecting characteristic sequential changes in EEG and EOG variables as wakefulness passes into sleep:

1. *alpha EEG,* generally continuous, with one or more *REMs* a few seconds prior to the "awakening";
2. *alpha EEG,* often discontinuous, with pronounced *SEMs* covering at least the 20–30 seconds of the record prior to the experimenter's call;
3. *descending Stage 1 EEG* (almost always, but not necessarily, accompanied by SEMs);
4. *descending Stage 2 EEG* of at least 30 seconds but not more than 2½ minutes duration (usually, but not necessarily, accompanied by SEMs).

It became apparent in pilot research that vagaries of individual sleep-cycle patterns over the course of an evening made it impractical to schedule awakenings in an order to which the experimenter was committed to rigorously adhere. The experimenters were, therefore, allowed freedom to schedule awakenings as they saw fit, with these restrictions: they were instructed to distribute awakenings for each category within and across sessions as evenly as possible, so that, for example, half of the alpha-SEM awakenings would occur early in the evening, half late in the evening, half in the first two sessions, half in the last two sessions. The experimenters were generally quite successful in achieving the desired intra- and intersession balance for all four categories of awakening.

The EEG and EOG record for each experimental session was independently scored by another experimenter than the experimenter assigned to that session. Rarely the two experimenters did not agree on the awakening category. An attempt was made, in such instances, to achieve reconciliation, which, when successful, might result in a change of the judgment made at the time of the experimental session; in a few instances, it proved impossible to achieve reconciliation, and these awakenings were excluded from all data analyses. For these reasons, the number of awakenings reported below per awakening category may sometimes deviate from the scheduled number of 6 per subject. Overall, the 9 subjects contributed 212 usable reports, 53 from alpha REM, 54 from alpha SEM, 47 from descending Stage 1, and 58 from descending Stage 2.

RESULTS

Incidence of Reportable Content

Employing the recall criterion of at least one item of specific *mental* (not necessarily *dream*) content reported, we computed percentages for each subject for each of the four EEG/EOG stages. The median of subject recall percentages for each of the four stages was 100%. Pooled recall percentages by stage were as follows: alpha REM, 96.2%; alpha SEM, 98.1%; Stage 1, 97.9%; Stage 2, 89.7%. Over the entire hypnagogic period, 202 of 212, or 95.3% of the total number of, reports contained at least one item of specific content.

EEG/EOG Stage and State of Consciousness

It will be recalled that subjects were asked, soon after being called by the experimenter, to judge their state of consciousness on a 5-point scale extending from "awake and alert" to "in deep sleep." Arbitrary numerical scale values (0–4) were assigned to the five alternatives supplied to the subject, with 0 assigned to "awake and alert," 1 to "awake but drowsy," etc. Mean values were then computed for each subject for each of the four EEG/EOG stages.

The medians of the subject means for each of these four stages are as follows (with the range of subject means given in parentheses): alpha REM, .80 (.17–2.17); alpha SEM, 1.50 (.50–2.67); Stage 1, 2.00 (1.25–2.67); Stage 2, 2.25 (1.67–2.67). These group averages reveal a decrease in wakefulness, an increase in drowsiness or sleepfulness, with successive EEG/EOG stages. Individual subject curves also reveal, almost without exception, a monotonic increasing relationship of drowsiness with successive EEG/EOG stages.

Although individual subject curves are all of the same general shape, they differ both in their absolute values and in their slope. The range of the subject mean values is large, particularly for alpha EEG stages, indicating rather wide individual differences in absolute judgments of state of consciousness for any one EEG/EOG stage. Subject 3's alpha-REM mean (2.17), for instance, exceeds the means of four subjects for Stage 2. Differences in slope reflect differences in the manner in which the different response alternatives were employed by subjects: some subjects tended to apply the same alternative to all awakenings, regardless of EEG/EOG stage (Subject 7, for instance, said that he was "drifting off to sleep" 19 of the 24 times that he was called by the experimenter), while other subjects spread their judgments much more evenly among the various alternatives. The identity of judgments of conscious state shown by some subjects in the face of orderly EEG/EOG variation suggests that the physiological variable may be superior to subjects' judgment in distinguishing shifts in the sequence of states of consciousness from wakefulness to sleep.

We examined the relationship between the early (first three) versus late (last three) "awakenings" made on each experimental night and judgments of state of consciousness to determine if time of night affected such judgments. Medians of subject judgments for early awakenings were: alpha REM, 1.00; alpha SEM, 1.00; Stage 1, 2.00; Stage 2, 2.00. Medians of the subject judgments for late awakenings were: alpha REM, .67; alpha SEM, 1.67; Stage 1, 2.00; Stage 2, 2.33. These data reveal little overall effect of time of night upon judgments of degree of wakefulness or sleep.

TABLE 1

INCIDENCE OF SENSORY IMAGERY BY EEG/EOG STAGE

	Alpha REM	Alpha SEM	Stage 1	Stage 2
modality				
Visual				
mdn.	100	86	100	100
range	(17–100)	(17–100)	(60–100)	(33–100)
M	78	81	89	83
Auditory				
mdn.	17	17	25	0
range	(0–33)	(0–50)	(0–100)	(0–100)
M	14	23	37	23
"Bodily"				
mdn.	20	29	0	17
range	(0–67)	(0–50)	(0–60)	(0–83)
M	25	32	22	23
Olfactory				
mdn.	0	0	0	0
range	(0)	(0–17)	(0)	(0–17)
M	0	4	0	2
Gustatory				
mdn.	0	0	0	0
range	(0)	(0)	(0)	(0)
M	0	0	0	0

Sensory Imagery

Each subject's answers to questions on the incidence of sensory imagery in various modalities were tabulated and percentages computed for each modality during each EEG/EOG stage. Table 1 presents the medians (and, in parentheses, the ranges) of these individual subject percentages and the mean percentage derived from a pooling of the data of all subjects. The data clearly reveal that hypnagogic experiences are primarily visual in character. Our "on-the-spot" study of the hypnagogic period thus fails to confirm McKellar's (1957) questionnaire results showing a predominance of auditory or auditory-plus-visual imagery while falling asleep. Part, but by no means all, of this difference may stem from our instructions that subjects carefully separate dreams and other experiences with conversational content "understood" to be occurring from those in which the subject actually experienced the auditory imagery of such conversations. It was fairly common for our subjects to report interpersonal verbal communication as dream content but to state that no "actual hearing" was involved. In any event, during the hypnagogic period we find auditory imagery, real *or* "understood," to be quite secondary to visual imagery, and about on a par with bodily (mostly kinaesthetic and tactile) imagery. Taste sensations were never reported and only 3 of 202 content-bearing reports contained descriptions of olfactory imagery.

The ranges presented in Table 1 imply that a wide diversity underlies the group averages for sensory imagery, and this is, in fact, the case. The high median values for visual imagery, for example, reflect the fact that 5 of our 9 subjects (3, 7, 8, 9, 10) are at, or one report from, a 100% level in all EEG/EOG stages. Of the other 4 subjects, two (4, 5) show an initially (alpha REM) low percentage of visual imagery (17% and 40%, respectively) with a steady rise to high values (86% and 100%) in Stage 2; one (1) hovers about a 50% level during all four EEG/EOG stages;

and one (11) starts at an initially (alpha REM) high value (100%) with a steady decline to a relatively low value (40%) in Stage 2. Similar diversity in imagery incidence and in imagery association with physiological stage was noted for auditory and bodily imagery.

The most intriguing feature of the imagery data, then, is their failure to demonstrate either any consistent association of imagery modality with particular EEG/EOG stage or any consistent change in the presence of imagery in any modality with successive EEG/EOG stages. This is but the first of several indications we have that individual differences in mental experience during the hypnagogic period are very large, probably far larger than those to be found in REM-period dreaming.

Affect

Percentages were computed for each subject of the incidence of experienced emotion in content-bearing reports from each of the four EEG/EOG stages. The medians of the subject affect percentages were as follows (with the ranges again presented in parentheses): alpha REM, 50% (0–83%); alpha SEM, 57% (0–100%); Stage 1, 33% (0–50%); Stage 2, 17% (0–50%). Individual subject curves showed much variability as to shape, but all subjects but one (5) reached their peak affect value during one of the two alpha stages. It is clear, then, that affect is at a reduced level during descending Stages 1 and 2 as compared with the preceding alpha phase. As we shall see below, this decline in affect is concomitant with an *increase* in hallucinatory dreamlike experience.

Where possible, a subclassification was made of the affirmative responses to the affect question according to whether the affect was predominantly positive (pleasurable feelings or positive feelings toward others) or negative (unpleasant feelings or hostile feelings toward others). Median subject percentages for positive affect were: alpha REM, 25%; alpha SEM, 29%; Stage 1, 17%; Stage 2, 0%. Median subject percentages for negative affect were: alpha REM, 17%; alpha SEM, 29%; Stage 1, 17%; Stage 2, 0%. Positive and negative affect were thus at about equivalent levels through all four EEG/EOG stages. These data do not indicate that hypnagogic experiences are predominantly unpleasant, as some writers (Critchley, 1955) have insisted. In fact, it is our impression that even in the minority of hypnagogic reports which do contain indications of the *presence* of affect it is quite rare for this affect to be at all *intense*. It might fairly be said that emotional flatness is the primary affective characteristic of the hypnagogic period, particularly of the dreamlike content which is reported from descending Stages 1 and 2. It is in this sense, perhaps, that Dement and Kleitman (1957a) and Oswald (1962) are most correct in asserting that hypnagogic experiences do not

TABLE 2

CONTROL OVER, AND REALITY-ORIENTATION OF, HYPNAGOGIC MENTATION

Question	''Yes'' replies by stage			
	Alpha REM	Alpha SEM	Stage 1	Stage 2
Were you controlling the course of your thoughts/imagery?				
mdn.	60	33	0	0
range	(14–100)	(0–100)	(0–33)	(0–33)
M	55	40	7	10
Were you aware that you were here in the laboratory?				
mdn.	67	33	17	0
range	(20–100)	(17–100)	(0–40)	(0–33)
M	67	49	13	8
Were you aware that you were observing the contents of your own mind?				
mdn.	67	50	20	17
range	(33–100)	(17–100)	(0–100)	(0–43)
M	71	53	26	20

seem quite "real" compared to nocturnal dreaming. Compare the present sleep-onset Stage 1 and 2 figures, for instance, with the finding of a previous study of REM-period dream content (Foulkes, 1960) that 50% of 88 REM-reports contained the subjective experience of some emotion.

Control and Reality Orientation

Upon tabulation of responses to questions concerning the degree of volitional control and the hallucinatory quality of mental activity during the hypnagogic period, percentages of affirmative response were computed for each subject for each EEG/EOG stage. Table 2 presents the median (and, in parentheses, the range) of individual subject percentages and the mean percentage derived from a pooling of the data of all subjects for each question for each of the four EEG/EOG categories. These data demonstrate a steady decline in volitional control and awareness of immediate circumstances and a steady rise in the frequency of hallucinatory experiences (felt to be "really" happening rather than simply the contents of one's own mind) with successive EEG/EOG stages. Almost without exception, individual subject curves for these three questions were of the same general shape as the group median curve: a monotonic decreasing relationship between "yes" replies and successive EEG/EOG stages. As was true with judgments of state of consciousness, however, the ranges in Table 2 indicate that subjects did differ fairly considerably in the absolute values inscribed by their curves.

Some subjects (3, 9), for instance, report a low frequency of volitional control even during alpha REM (17% and 14%, respectively), with its total loss during descending Stages 1 and 2. Other subjects (1, 4) report a complete retention of such control during alpha REM and complete or almost complete retention of such control during alpha SEM (67% and 100%, respectively), and although they report a substantial loss of volitional control following this point, they continue to report some retention thereof during both descending Stages 1 and 2. Similar individual differences were observed with regard to judgments of orientation to time and place and judgments of hallucinatory quality of mentation.

Not unexpectedly, responses to these three questions tended to be correlated. Compared with subjects who report a substantial retention of volitional control during the early, alpha EEG stages, subjects who report a relatively early (e.g., alpha REM and alpha SEM) loss of volitional control also report an earlier loss of awareness of immediate circumstances (i.e., the lab setting) and an earlier loss of the purely mental quality of the events which they are experiencing. As the data in Table 2 show, loss of volitional control tends to occur first, then a loss of awareness of immediate environment, and finally a loss of the sense that one's experience is purely mental rather than actually transpiring out in the "real" world.

In general, subjects 3, 7, 8, 9, and 10 showed signs of considerably reduced control and reality-orientation during alpha stages, while subjects 1, 4, and 5 retained a relatively high degree of control and reality-orientation in these stages and only later, during descending Stages 1 and 2, demonstrated a comparable reduction in these ego functions. MMPIs were collected from the subjects, but in this small sample of subjects they revealed no striking clues to personality differences associated with these different performances during the hypnagogic period. Our clinical impressions, gained from observation of subjects in the laboratory, were, however, that subjects with earlier signs of the loss of ego-control and reality-testing were less anxious and constricted than those who "held on" to these functions during the alpha phases. Their earlier engagement in regressive thought processes seems to us to indicate a greater willingness and ability to explore and accept their own fantasies. Future research, however, might well focus upon the personality variables associated with such differences in ego functioning during the hypnagogic period to attempt to better substantiate these impressions.

A comparison of reports from "awakenings" coming "early" and "late" in the night indicated that time of night is one variable determining the extent of loss of reality-testing. For example, the medians of the subject proportions of nonhallucinatory reports early in the night are: alpha REM, 100%; alpha SEM, 67%; Stage 1, 25%; Stage 2, 33%—

while for late in the night the comparable proportions are 50%, 50%, 0%, and 0%. Subject 3, moreover, did not have a single nonhallucinatory report late in the night, while 50% of her reports early in the night were nonhallucinatory. The group data show, however, that there are still substantial differences in degree of reality-testing between alpha phases and descending Stages 1 and 2 quite apart from the influence of the time of night variable.

We also examined the relationship of amount of time spent in bed since either lights-out or the termination of the prior interview with the hallucinatory quality of reported mentation. No consistent relationship existed between these two variables. For example, both Subjects 1 and 4 had relatively few hallucinatory reports in any EEG/EOG phase: Subject 1 had the shortest median pre-"awakening" time in our sample (4.5 minutes), while Subject 4 had the longest time (23.0 minutes). It does not seem, then, that time in bed prior to the experimenter's call is a determinant of the level of reality-testing reported by the subject upon being called.

General Nature of Content

In an attempt to provide suitable experimental data for resolving the issue of the hypnagogic dream, we tabulated the percentage of content-producing awakenings on which a *dreamlike* report was given (i.e., a dramatic episode which, for the most part, at least, to the subject seemed to be "really" happening[2]) for each subject for each EEG/EOG stage of the hypnagogic period. These data are presented in Table 3. It will be noted that "dreams," that is, hallucinated dramatic episodes, did occur with considerable frequency during the hypnagogic period, most frequently during descending Stages 1 and 2.

TABLE 3

INCIDENCE OF HALLUCINATORY, DRAMATIC EPISODES ("DREAMS")

Subject	Percentage of content-bearing reports by stage			
	Alpha REM	Alpha SEM	Stage 1	Stage 2
1	0	33	40	17
3	83	67	100	100
4	0	0	75	57
5	0	0	25	20
7	50	50	100	100
8	33	67	100	100
9	29	83	100	80
10	50	67	60	100
11	20	29	67	60
Mdn. of *S*%s	29	50	75	80
Pooled *M*	31	43	76	71

But dreams were also reported during those periods of more or less continuous alpha which have been heretofore identified with wakefulness and with wakeful levels of mental functioning. That these were something other than engrossing waking fantasies is indicated by the following facts: (*a*) by definition, these experiences had to partake of an hallucinatory quality which is lacking in even the most engrossing of waking fantasies; (*b*) subjects generally described such experiences as more like night-dreams than any form of waking fantasy; (*c*) subjects described their state prior to awakenings producing such content as "drowsy," "drifting off to sleep," or "light sleep"; and (*d*) such content showed distortions and symbolic transformations of a type generally assumed to be absent in voluntary daytime fantasy. To illustrate these four points, let us consider one of Subject 3's alpha-REM "dreams." She reported that she saw her husband open the door to a room in a house which she had once visited in her own hometown.

> I think he was looking for an ironing board. . . . I really didn't think he was looking for an

---

[2] This definition of a dream seems to us to incorporate the principal features commonly felt to be characteristic of the phenomenon. We reserve for later a discussion of the sense in which hypnagogic dreams are similar to, and different from, the REM-period hallucination which is, at present, the most usual referent of the term "dream." It might be wondered at this point, however, how subjects themselves characterized the content of their reports. In answering the interview question directed at this point, subjects would generally agree with the classification produced by our criteria in the sense that they would feel that hallucinatory or quasi-hallucinatory, dramatic content was more like a night-dream than like a daydream, memory, or waking thought. In addition, some content which was not, by our criteria, dream content *would* be so classified by subjects themselves. Individual variation in criteria underlying these judgments, however, seemed to indicate that it would be preferable to rely primarily upon a set of explicit criteria applied uniformly across subjects.

ironing board, I just knew he was looking for a room, but when he opened the door of this room and went into this room that I had been in before, I saw the ironing board and I thought that was what he was looking for. It's very confusing. It seems to me it was very important what he was looking for, seems like there was kind of a strain on me, too, that I was helping him. But when he opened the door, instead of going in and taking it, he just closed the door again. This was kind of strange. . . .

The subject described this experience as "definitely" similar to a night-dream ("because it was incongruous events thrown together"), as completely hallucinatory, and as having transpired during "light sleep."

Oswald (1962) feels that hypnagogic hallucinations differ from "true" dreams in that subjects do not hallucinate their own participation in their experiences at sleep onset. With this suggestion in mind, we examined the content of the "dreams" reported by our subjects during the hypnagogic period and their answers to the question "Were you an active participant in what you experienced, or just passively observing it?" It was found that in a clear majority of these dreams, subjects *did* hallucinate their own active participation, and that the frequency of such hallucinated participation increased with the succession of EEG/EOG stages. Pooled data from all subjects showed hallucinated subject-participation in 62% of alpha-REM dreams, 78% of alpha-SEM dreams, 83% of Stage 1 dreams, and 97% of Stage 2 dreams. It appears, then, that hypnagogic dreams can, and do, have that "living in" quality which Oswald has reserved for the REM-period dream.

It should not be assumed from the foregoing evidence that the self is generally the only character in a typical hypnagogic dream. Considering the pooled data of all subjects, of those reports classified as dreams, 69% from alpha REM, 83% from alpha SEM, 100% from Stage 1, and 81% from Stage 2 contained, or made explicit reference to, some human character other than the subject himself.

The intersubject variability revealed in Table 3 is extensive, with Subjects 3, 7, 8, 9, and 10 being early (alpha EEG) and frequent dreamers, while Subjects 1, 4, 5, and 11 are "later" and less frequent dreamers.

DISCUSSION

Hypnagogic "Dreams"

By one definition of "dream"—an hallucinatory, dramatic episode—dreams occur frequently during the hypnagogic period, particularly during descending Stages 1 and 2. We must now inquire whether, or in what sense, these hypnagogic dreams are similar to nocturnal dreams occurring during REM periods.

While REM-period dreams can be, and often are, quite fantastic and distorted, they are still usually tightly organized and internally coherent. That hypnagogic hallucinations can be fantastic and distorted has generally been acknowledged in the literature, but it is usually implied that they are not very well organized as compared with nocturnal dreams—that they are fleeting and disordered sensations (e.g., Dement & Kleitman, 1957a). Our data, however, contain numerous examples of the fact that hypnagogic hallucinations can be both fantastic *and* well-organized, as well-organized as any REM-period dream. To be sure, some hypnagogic experiences are fleeting progressions of visual imagery or dissociated images and thoughts. And, even some hypnagogic dreams show traces, in the form of dissociated elements of visual imagery, of impairment of the ego's synthetic function. But the typical hypnagogic dream seems no less well-organized than its REM-period counterpart.

The REM-period dream has also seemed unique (compared, for instance, with non-REM mentation) in the extent to which it is "primary-process," that is, the presence of displacement and condensation make its images bizarre and symbolic so that they are susceptible to depth-level psychoanalytic interpretation. Many of the present sample of hypnagogic reports, both dream and non-dream, seem to be equally revealing, in this sense, as any REM-period dream, and much more revealing than typical non-REM content. Consider, for instance, the following two dream reports, collected from Subject 8 during the course of his third experimental night:

5. I was sort of standing on the edge of a pool table, but instead of pool balls rolling around out there, they were sort of balloons, half blown up with water in them. The balloons were rolling around on

> the table. I was just seeing the area of the table. I made some remark, something like "It'll never come back" or something like that. I think I felt sort of a longing feeling, I don't know for what, though. Those balloons were yellow and they were lopsided. They weren't round like pool balls, but somehow they rolled. They shouldn't have really, they rolled well but they weren't in the shape of good rollering. They sort of looked like pears.
> (Stage 2)
>
> 6. I was looking at sort of a low lazy-Susan type of thing which was on the floor under a typewriter stand. I was in a very peculiar vantage point. I was looking down between the legs there, and never got above this. It was made out of crystal, and it was a platter type of arrangement. In the middle there was a stem and a little ball on the top, and I first saw there was blood in the little glass thing. In the middle it was full of blood; and then as it developed, the blood turned into what looked like sort of cocktail sauce that's served with shrimp; and then as it developed more, little shrimp started appearing around the little glass where the sauce was; and then as it developed more, more shrimp appeared on the platter below; and just as you called, a dog was walking over there and was just about ready to help himself to a few of the shrimp. This was mostly in vague symbols because there was nothing realistic about this at all. The dog was strange too, because his head, you know where his snout comes down to his nose, this part was all, like it was sawed off. In other words, his face was completely squared off and he didn't have any nose. I remember when I saw the dog coming out, I looked at him, made me feel uneasy. I think I know what that meant. [*E:* What?] Well, I just wondered whether Freud might interpret that as a castration complex or something, I don't know.
> (Alpha SEM)

In the second report the subject himself provides the key to at least one possible interpretative theme which can be applied to both dreams. Whatever particular form the interpretation of these two dreams might take, however, it is clear that the content is readily susceptible to the same general kinds of interpretation employed with nocturnal dream content. In Freudian terms, hypnagogic content, particularly hypnagogic dream content, *can* be as regressive, symbolic, etc. as REM-period dream content. We are not able to agree with Kubie (1943), then, that hypnagogic dreaming must be much more limited in intent and scope than true nocturnal dreaming.

With respect to degree of mental organization and level of mental functioning, then, hypnagogic "dreams" can be just that, *dreams*. Dreaming thus occurs in the absence of REMs, emergent Stage 1, or any kind of Stage 1 altogether (a finding foreshadowed by Kamiya's [1961] report that when his subjects felt that they were asleep prior to hypnagogic Stage 2 awakenings, they invariably reported dreaming). From this finding there stem a number of important implications for the study of dreaming. Some of these are at a practical level. For instance: the number of dreams a subject experiences cannot be gauged solely from the number or nature of his REM periods; all night recording is not necessary to gain clinically meaningful dream series; etc. Other implications are at a more theoretical level.

First, the REM period cannot be so much the unique state of existence as has been heretofore imagined (Dement, in press; Snyder, 1963) nor can the elucidation of the physiological mechanisms underlying the REM period provide any kind of completely satisfactory answer to questions concerning the psychophysiology of dreaming. Without meaning to deny some important differences—to be discussed below—between hypnagogic and REM-period dreams, *the similarities are sufficiently crucial for us to insist that an adequate account of psychophysiology of dreaming must explain the occurrence of dreaming in the absence, as well as in the presence, of REMs and during phases of continuous alpha and descending Stage 1 and 2 EEG patterns as well as during emergent EEG Stage 1.*

Secondly, it will be necessary to re-examine many of the conclusions which have been drawn from experimental evidence derived solely from investigation of REM periods. Much of the "definitive" data which EEG dream researchers have successfully diffused to professional colleagues and the general public may apply to only one species of dreaming and not to the entire class. Conversely, many of the everyday notions about dreaming which were hastily discarded in the wake of REM-period research may turn out to be valid for at least some kinds of dreaming, those which do *not* occur during Stage 1 REM periods. To illustrate some of the possibilities of the kind of re-examination which we are proposing, let us now inquire into possible *differences* between hypnagogic and REM-period dreams.

Folklore has had it that dreams occur instantaneously, and the older dream literature is full of reports of dreams which seem to compress the usual time sequences of waking life, for example, the case of Maury's guillotine dream (Maury, 1878). REM-period research has assured us, however, that the course of time of dream happenings accurately mirrors that of the corresponding events in waking life. REM periods, it was noted, are rarely less than 10 minutes long, sometimes last an hour or more. That this time is devoted to dreaming one dream or one series of related dream episodes was suggested by several lines of evidence. Dement and Kleitman (1957b) showed that subjects can judge, with a high degree of accuracy, whether the experimenter has allowed them an undisturbed REM period of 5 or 15 minutes duration prior to an awakening on the basis of the extent of their dream recall. Dement (1958) also showed that the acting out of REM-period dreams takes about as long as the duration of the corresponding REM period, and stimulus-incorporation data (Dement & Wolpert, 1958) also support the idea that the course of time in the REM period is identical to that in waking life.

Our data suggest, however, that *some* hypnagogic dreams occur, if not instantaneously, at least much more quickly than the events they represent. The mechanism of this temporal condensation is the omission of visual continuity, so that the subject reports that his visual imagery was something more like a succession of snapshots than like a movie.

One of Subject 10's alpha-REM dreams illustrates this mechanism of a rapid snapshotlike sequence which is still experienced as a dream, the continuity residing in the dreamer's "understanding" rather than in his visual imagery. His dream was of driving from California to Tijuana, Mexico, of entering a bar there which featured a dancing girl, and of then leaving the bar in the company of this girl. In discussing this dream, the subject commented that he had really had only three brief images—an image of himself driving at one particular place in California, an image of the inside of the bar, and an image of the girl walking with him in the street. The dream as reported, and apparently as experienced, had some continuity, but not the visual, movielike continuity which is most typical of REM-period dreams. It seems to have been more like three slides with a great deal of "filling in" or secondary revision taking place at a nonvisual level, allowing the dreamer to condense a relatively long action-sequence into a brief visual presentation.

Another subject (9) characterized one of his hypnagogic Stage 2 dreams as follows:

> There was a connection through thought. The images themselves were like a series of stills, but it seemed to be unity through the thought itself.

It has been reported that certain episodes of true nocturnal dreams may occasionally take this visually "disconnected" form (Roffwarg et al., 1962). In the hypnagogic period, however, we find that entire dreams may be constructed in this manner, that is, there is a total absence of visual continuity. The mode of construction of such hypnagogic dreams may well be the factor giving rise to the notion that dreams occur instantaneously. Such a notion may not be entirely wrong, as we have recently been led to believe. It may well apply at least somewhat accurately to some members of one class of dreams, those occurring at sleep onset, even if it seems to be not at all supported by the evidence with respect to another class of dreams, those occurring through the night during periods of emergent EEG Stage 1.

It has also been concluded from the study of REM-period dreams that there are no substantial individual differences in the occurrence of dreaming. For example, people who describe themselves as chronic nondreamers, say they can hardly ever recall dreaming, do not have fewer REM periods than do self-described chronic dreamers, those who seem to be able to recall a dream or more almost every night of the week (Antrobus, 1962; Goodenough, Shapiro, Holden, & Steinschriber, 1959). Moreover, when such persons are awakened in the laboratory from REM periods, they do recall dreams (Goodenough et al., 1959). These REM-period findings have led to the conclusion that there are recallers and nonrecallers, but no such thing as dreamers and nondreamers.

In the present study, however, there are rather wide individual differences in the dreamlike quality of content, and in the association of such content with particular EEG and EOG stages. Some of our subjects (e.g., 3) are more or less chronic dreamers, dream throughout the several different EEG/EOG stages of the hypnagogic period, while others (e.g., 5) hardly dream at all and others (e.g., 4) dream with any substantial frenquency only during certain EEG/EOG stages. The large individual differences in the immediate recall of dreams during the hypnagogic period provide an obviously fertile field for research on the relationship of personality characteristics to dream occurrence. The results of such studies may well be much more illuminating than any comparable research on REM-period dreams. And, "nonrecallers" may turn out to be hypnagogic nondreamers.

The examples of temporal relationships and individual differences are perhaps sufficient to establish the point—that many of the popular and previous scientific findings on dreaming which have been discarded in the light of REM-period research may find confirmation in research on hypnagogic dreaming. There may be at least two classes of dreams. If so, the validation or rejection of hypotheses and previous research concerning dreaming will have to proceed separately for each of these classes.

Perhaps one of the clearest differences between a typical hypnagogic report and one from a relatively long REM period lies in the amount of substantive content contained therein. In the present study we obtained no reports of the length (e.g., 2–3 single space typed pages) sometimes obtained from REM periods. In accordance with the subjects' impressions of brevity of hypnagogic experiences, hypnagogic dream reports were generally fairly short. This difference between hypnagogic and full-term REM-period dream reports is a real one, and it provides some justification for applying the label "dreamlet" (Foulkes, 1962) to hallucinatory hypnagogic content.

It would be a mistake, however, to draw any sharp distinction between REM-period dreaming and hypnagogic dreaming simply on the basis of length of reported content. First of all, the present sample of hypnagogic reports does not differ very strikingly with respect to length from reports obtained from awakenings made during the earlier minutes of REM periods or following the completion of brief, that is, early night, REM periods. Also, some of these hypnagogic reports do manage to include a considerable amount of material. For example, Subject 11 reported the following dream from an alpha-SEM awakening:

> Well, it's kind of a long story. As to where you interrupted me, I'll explain. Anyway, I got out of here [the laboratory], and I went back down to the pub, and she [a girl subject is interested in courting, although she has indicated to him that she already has a boyfriend] was getting off work, and I said, why don't you ditch this boyfriend that's supposed to pick you up and hop on a subway train and go up to my place and have a drink. She's interested in interior decorating, so I said, I'll show you my place and show you what I'm doing to it. And so she said OK, and a few choice lines came in between, and she said, at one time or another, I'm in love with the guy. And I said, well, honey, I fell in love with you the first time I saw you, and this and that. Then trotting down the steps to the subway, and she went through one gate and I went through another, and then down the platform. She was standing there, and I lit a cigarette for her, like we'd just met, you know, so in case the guy saw us, or something. . . . Then when the right train came, I'd start to get on, and she got on, and then that was all there. The next thing we were in my apartment, just coming in the door and closed the door, had a few drinks, next thing I was in bed with her, spent the night. [*E:* Sexual intercourse?] Yes, I think I was imagining myself in that position. The next thing I remembered was I was thinking about making a phone call and then I remembered I couldn't make a phone call from downstairs because I didn't have one of my own. I called up sick for work, and I spent the day, so we got pretty thick, see (laughs), and at the time you called me, I was saying, you really want to get rid of this guy, the way to do it, to get him really off your back is if sometime he gets rough and hits you, you can have him arrested. And that's where you interrupted me.

There must be many other important differences between hypnagogic and REM-period dreams. Surely it must make a considerable difference whether a dream develops out of relaxed wakefulness or from an hour or more of sleep. In particular, we may wonder about possible differences in the functional significance of the two classes of dreams.

Incidence of Particular Forms of Hypnagogic Experience

One of the secondary purposes of the present study was the determination of the psychophysiological state associated with, and of the frequency of occurrence of, certain unusual experiences said to be characteristic of the hypnagogic period, e.g., the Isakower (1938) phenomenon (peculiar bodily distortions, often involving the mouth), the autosymbolic effect (Silberer, 1951), and experiences of falling (McKellar, 1957). It is our impression, based upon the data of the present study, that such experiences occur far less frequently than previously published accounts imply. The frequency of occurrence of such experiences is most likely over-estimated by virtue of unrepresentative sampling—these unusual hypnagogic experiences may be related to "startle" or momentary arousal, whereas other, more frequent hypnagogic experiences go relatively unnoticed because they are less unusual and because they are not associated with any disturbance of the orderly progression from wakefulness to sleep.

Five of our 9 subjects *never* reported distortions of bodily representation; these subjects accounted for a total of 118 experimental interviews. The remaining 4 subjects produced a total of 5 instances of peculiar feelings of bodily weightlessness, 6 instances of distortions of overall body size, and 1 instance of distortion in the size of a bodily part. In all of the hallucinated changes of body size, the change was *seen*, rather than sensed in any more direct manner. Only one of the 202 content-bearing reports contained a clear-cut instance of the Isakower phenomenon, a directly-sensed bodily distortion, and, as was often true of Isakower's examples, it was oral in character. This extremely low incidence of Isakowerlike hallucinated body transformations is especially striking in view of the fact that each interview contained a question directed specifically toward such content, a question undoubtedly suggesting that such content might be occurring during the hypnagogic period.

The incidence of experiences of "falling" was likewise extremely low: one case each for Subjects 3 and 4, both from Stage 1. This represents less than 1% of our experimental trials.

Silberer's "autosymbolic pheomenon," that is, the spontaneous process during the hypnagogic state in which abstract thoughts, bodily feelings, or psychic states become transformed into concrete perceptions, is difficult to substantiate directly from our data. Although we found only a few reports suggesting its occurrence, we did not obtain the second of the two conditions which Silberer felt necessary for its occurrence: drowsiness *and* an effort to think.

CONCLUSIONS

The present data do not challenge the validity of any of the now extensive observations associating nocturnal dreaming with REM sleep. They do, however, suggest that mentation often sufficiently dreamlike (visual, dramatic, hallucinatory, and often distorted) to justify the label "dreaming" can and does occur, in the absence of REMs or ascending Stage 1, at sleep onset. The dreamlike properties of hypnagogic mentation are, in particular, much more striking than those of non-REM mentation, which has proven, from the very outset, to be readily distinguishable from REM dreaming.

We have attempted to demonstrate that a number of the suggestions of previous researchers (and, more often, anecdotal observers) as to features differentiating hypnagogic dreaming from true nocturnal dreaming are not borne out by our data. To be specific: in contrast to these earlier publications, our data indicate that hypnagogic experiences are often believed to represent experiences actually taking place in the "real world" (cf. Dement & Kleitman, 1957a); they often do involve the self in an active role (cf. Oswald, 1962); they are most often visual (cf. McKellar, 1957) and lacking in either positive or negative affect (cf. Critchley, 1955); they are most often internally coherent (cf. Dement & Kleitman, 1957a); they can be as symbolic and primary-process as the true nocturnal dream (cf. Kubie, 1943).

While, in these various respects, hypnagogic dreaming does not prove to be easily distinguishable from REM-period dreaming, we have suggested several other areas in which attempts at psychological differentiation of hypnagogic dreams from REM-period dreams may prove more rewarding (e.g., mode of construction, susceptibility to individual variation, psychodynamic function). There is already, of course, the obvious factor of physiological differentiation of the two classes of dreams—ascending EEG Stage 1 versus descending EEG Stage 1 *or* alpha EEG *or* descending EEG Stage 2, and REMs versus SEMs or no eye movements at all.

Whatever the outcome of future attempts at discriminating the content of REM dreams from the content of hypnagogic dreams, it is the psychological similarity that must at present claim our attention. With the finding that dreamlike mental activity may occur in several organismic states, it would now seem best to question generalizations about dreaming which have been made from data collected in only one of the dream states, *viz.*, in the emergent EEG Stage 1 REM period.

REFERENCES

Antrobus, Judith S. Patterns of dreaming and dream recall. (Doctoral dissertation, Columbia University) Ann Arbor, Mich.: University Microfilms, 1963, No. 63-6101.

Archer, W. "Hypnagogic illusions," or dozing-off dreams. In R. L. Woods (Ed.), *The world of dreams*. New York: Random House, 1947. Pp. 753–760.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. *Journal of Applied Physiology,* 1955, **8**, 1–10.

Critchley, M. The pre-dormitum. *Revue Neurologique,* 1955, **93**, 101–106.

Dement, W. C. The physiology of dreaming. Unpublished doctoral dissertation, University of Chicago, 1958.

Dement, W. C. The psychophysiology of dreaming. In *The symposium, le rêve et les societés humaines,* in press.

Dement, W. C., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (a)

Dement, W. C., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (b)

Dement, W. C., & Wolpert, E. The relation of eye movements, body motility, and external stimuli to dreaming. *Journal of Experimental Psychology,* 1958, **55**, 543–553.

Foulkes, W. D. Dream reports from different stages of sleep. Unpublished doctoral dissertation, University of Chicago, 1960.

Foulkes, W. D. Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology,* 1962, **65**, 14–25.

Goodenough, D. R., Shapiro, A., Holden, M., & Steinschriber, L. A comparison of "dreamers" and "non-dreamers": Eye movements, electroencephalograms, and the recall of dreams. *Journal of Abnormal and Social Psychology,* 1959, **59**, 259–302.

Isakower, O. A contribution to the patho-psychology of phenomena associated with falling asleep. *International Journal of Psychoanalysis,* 1938, **19**, 331–345.

Kamiya, J. Behavioral, subjective, and physiological aspects of drowsiness and sleep. In D. W. Fiske and S. R. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey Press, 1961. Pp. 145–174.

Kubie, L. The use of induced hypnagogic reveries in the recovery of repressed amnesic data. *Bulletin of the Menninger Clinic,* 1943, **7**, 172–182.

McKellar, P. *Imagination and thinking*. New York: Basic Books, 1957.

Maron, Louise, Rechtschaffen, A., & Wolpert, E. A. The sleep cycle during napping. *Archives of General Psychiatry,* 1964, **11**, 503–508.

Maury, L.-F. A. *Le sommeil et les rêves*. Paris: Didier, 1878.

Monroe, L. J., Rechtschaffen, A., Foulkes, D., & Jensen, Judith. The discriminability of REM and NREM reports. *Journal of Personality and Social Psychology,* in press.

Oswald, I. *Sleeping and waking: Physiology and psychology*. Amsterdam, Netherlands: Elsevier, 1962.

Rechtschaffen, A., Verdone, P., & Wheaton, Joy. Reports of mental activity during sleep. *Canadian Psychiatric Association Journal,* 1963, **8**, 409–414.

Rechtschaffen, A., Wolpert, E. A., Dement, W. C., Mitchell, S. A., & Fisher, C. Nocturnal sleep of narcoleptics. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 599–609.

Roffwarg, H. P., Dement, W. C., Muzio, J. N., & Fisher, C. Dream imagery: Relationship to rapid eye movements of sleep. *Archives of General Psychiatry,* 1962, **7**, 235–258.

Silberer, H. Report on a method of eliciting and observing certain symbolic hallucination-phenomena. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia University, Bureau of Publications, 1951. Pp. 195–207.

Slight, D. Hypnagogic phenomena. *Journal of Abnormal and Social Psychology,* 1924, **19**, 274–282.

Snyder, F. The new biology of dreaming. *Archives of General Psychiatry,* 1963, **8**, 381–391.

Vogel, G. Studies in psychophysiology of dreams: III. The dream of narcolepsy. *Archives of General Psychiatry,* 1960, **3**, 421–428.

Williams, H. L., Tepas, D. I., & Morlock, H. C. Evoked responses to clicks and electroencephalographic stages of sleep in man. *Science,* 1962, **138**, 685–686.

(Received December 7, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 244–245

PROVERB WORD COUNTS AS A MEASURE OF OVERINCLUSIVENESS IN DELUSIONAL SCHIZOPHRENICS

ROBERT H. GOLDSTEIN AND LEONARD F. SALZMAN
University of Rochester, School of Medicine and Dentistry, Rochester, New York

A partial replication was conducted of Payne's (1964) report concerning the relationship between delusions and overinclusive thinking as measured by proverbs word counts. Word counts were made of Gorham Proverbs Test responses given by age- and intelligence-matched groups of 21 delusional schizophrenics, 21 nondelusional schizophrenics, and 21 nonschizophrenic, nondelusional psychiatric in-patients. 23 paranoid, 23 nonparanoid, and 23 nonschizophrenic control Ss, similarly matched, from the same population, were also compared. No significant differences in proverb word counts were found among the groups. It was concluded that, although the possibility of increased irrelevant side remarks could not be excluded, paranoid or delusional schizophrenics do not produce higher proverb relevant word counts.

In a recent paper Payne, Caird, and Laverty (1964) have attempted to interpret the presence of delusional thinking in terms of the concept of overinclusiveness. They have suggested that overinclusive thinking results in "unwarranted generalizations" and an inability to screen out irrelevant perceptions. They have further suggested that this kind of perceptual overinclusion may play a role in the development of delusional thinking. In support of this hypothesis, they presented data relating the presence of delusions in a psychiatric population to overinclusiveness in proverb interpretation, as defined by the "average number of words required to explain each proverb."

As part of a larger study of conceptual thinking in psychiatric patients, it was possible to attempt a partial replication of these results by an analysis of Gorham Proverbs Test responses given by a group of schizophrenic and nonschizophrenic psychiatric patients. Our subjects consisted of patients admitted to an acute in-patient psychiatric unit in a large university teaching hospital. Gorham Proverbs (Gorham, 1956) were administered orally on an individual basis within 48 hours of admission and before any systematic drug therapy had been instituted. Proverb interpretations were recorded verbatim by the examiners, none of whom were familiar with the hypothesis proposed by Payne et al. Subsequently, a word count was performed on these data following as closely as possible the criteria described by Payne. The data were analyzed in two ways.

First, the schizophrenic group was partitioned into *delusional* and *nondelusional* groups based upon a detailed examination of each patient's chart. Patient's charts, which were examined following their discharge, contained complete history, mental status examination, and records of hospital course. Each of the delusional, nondelusional, and control groups contained 8 males and 13 females. The groups were matched on the basis of age and Shipley-Hartford Vocabulary scores before word counts were examined.

Because Payne's reasoning also led to the same hypothesis with respect to paranoid

TABLE 1

AGE, VOCABULARY, AND WORD COUNT SCORES BY DELUSIONAL STATUS

		Deluded schizophrenics	Nondeluded schizophrenics	Controls
	N	21	21	21
Age				
	M	34.81	31.95	32.33
	σ	10.34	10.21	8.88
		F = .50		
Vocabulary				
	M	28.33	27.86	28.90
	σ	6.77	6.89	5.59
		F = .13		
Word count				
	M	10.34	11.60	10.49
	σ	3.29	2.81	3.17
		F = 1.00		

TABLE 2

AGE, VOCABULARY, AND WORD COUNT SCORES BY DIAGNOSIS

	Paranoid schizophrenics	Nonparanoid schizophrenics	Controls
N	23	23	23
Age			
M	34.52	32.65	33.22
σ	10.20	11.34	10.24
	F = .19		
Vocabulary			
M	27.61	27.39	28.04
σ	7.34	7.38	5.30
	F = .05		
Word count			
M	11.28	10.71	11.29
σ	3.23	3.46	3.83
	F = .84		

versus nonparanoid schizophrenics, the schizophrenics were also divided into *paranoid* and *nonparanoid* diagnostic groups. The paranoid, nonparanoid, and control groups each contained 7 males and 16 females similarly matched on age and vocabulary scores. No delusional or paranoid patients were present in the nonschizophrenic control group, which consisted of neurotics and character disorders. There were no significant differences among the groups in terms of age or general intellectual level, as measured by the Shipley-Hartford Vocabulary Scale (see Tables 1 and 2). All word counts were carried out without knowledge of the patient's diagnosis or delusional status.

Tables 1 and 2 present the mean number of words per proverb interpretation in each group. Analysis of variance indicated no significant differences among any of the patient groups, either on the basis of the *delusional* versus *nondelusional* dichotomy or the *paranoid* versus *nonparanoid* schizophrenic grouping.

Our data could be interpreted as not confirming Payne's hypothesis of a relationship between overinclusive thinking and delusional ideation. However, it should be noted that Payne's data were gathered from tape-recorded responses, while ours derive from verbatim written records. This may account for the higher mean word counts found in Payne's population. It is possible that our examiners did not record completely irrelevant comments introduced into the patient's proverb interpretation, in spite of instructions for *completely* verbatim recording. In this case, one might conclude that Payne's data may reflect primarily a greater frequency of irrelevant side remarks not germane to the interpretation of the proverb per se.

Our own data do not, however, support Payne's contention that paranoid or delusional patients would manifest "a more extensive explanation," or "a more complex and extensive concept" in their proverb relevant responses.

REFERENCES

GORHAM, D. R. A proverbs test for clinical and experimental use. *Psychological Reports,* 1956, 2 (Monograph Supplement No. 1), 1–12.

PAYNE, R. W., CAIRD, W. K., & LAVERTY, S. G. Overinclusive thinking and delusions in schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1964, **68**, 562–566.

(Early publication received December 30, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 246–250

SOME DETERMINANTS OF PERSISTENCE IN CHRONIC SCHIZOPHRENIC SUBJECTS

RICHARD BLUMENTHAL, JULIAN MELTZOFF
Veterans Administration Outpatient Clinic Brooklyn, New York
AND SELIG ROSENBERG
Veterans Administration Hospital Northport, Long Island

This investigation of persistence in schizophrenics on tasks differing in complexity and interest under ego-involved and task-oriented conditions utilized 3 groups of male schizophrenics: 30 well-adjusted and 30 poorly-adjusted outpatients, and 30 regressed hospitalized patients. Each S worked at 1 of 3 tasks representing low, moderate, and high degrees of interest and complexity. Half of each group received ego-involved and half received task-oriented instructions. Outpatient schizophrenics were clearly superior to hospitalized patients in persistence. Persistence was greater under ego-involved than under task-oriented conditions, regardless of task or level of psychopathology, but there was no difference in persistence scores across tasks.

An exceptionally low level of drive, pervasive in all aspects of behavior, is characteristic of chronic schizophrenics. Typically these patients appear unable to persist at goal-directed activities over sustained periods of time. More often than not, they have been observed to withdraw from the initiation of such activities or interrupt those once begun. Such minimal interaction with the environment isolates them from the social world, and contributes to the maintenance of their illness by preventing them from making a start towards the restoration of their productive functions.

Although the study of persistence has had a long history in psychology (Eysenck, 1953; Lewin, 1935; Ryans, 1939) there have been almost no systematic studies of persistence in schizophrenia. Studies of the effects of motivational variables on cognitive and perceptual processes have focused on these processes rather than on the motivational variables themselves.

In general, factorial studies with normals have shown that persistence is a well-defined trait that can be subsumed under such achievement motivation factors as endurance and ambition (Eysenck, 1953; Guilford, 1959). Experimental studies (Lewin, 1935) with normals have demonstrated that persistence is systematically related to selected personality characteristics of the individual as well as the structure of the task. In schizophrenic subjects as well, one would expect to find persistence to be a well-defined trait that varies systematically in relation to personality characteristics and task structure. A growing body of experimental literature on the behavior of schizophrenics suggests a number of relevant personality and situational variables that should be systematically related to the presence of persistence in these subjects.

One important variable receiving increasing attention is the subjects's degree of psychotic disturbance. It has been demonstrated (Blumenthal, 1964a; Hunter, 1960; Schooler & Spohn, 1960) that regressed schizophrenics show marked performance deficits in several behavior modalities in a variety of experimental situations as compared with subjects in whom there is only a moderate degree of psychotic disturbance. These studies have generally confined themselves to hospitalized patients. Ambulatory schizophrenics, like their hospitalized counterparts, can be placed on a continuum of psychotic disturbance ranging from those who are barely able to stay outside the hospital, to those who can function effectively in the community. On the basis of the research previously cited, it seems reasonable to hypothesize a relationship between persistence and extent of psychopathology.

An additional and equally relevant determinant of the patients' performance is their perception of the experimental situation. One of the more fruitful concepts used in examining this variable is that of ego involvement as against task-orientation. In a review of a large body of experimental literature with normals, Iverson and Reuder (1956) have reported that moderate levels of ego involvement facilitate performance on a variety of tasks. Since schizophrenics are quite concerned with their self-concept and self-esteem, it should follow that experimental conditions that arouse these feelings would have a marked effect on their performance.

A situational variable logically related to persistence is the structure of the task that the subject performs. Of the many aspects of task structure subject to analysis, interest and complexity are especially relevant to the arousal of persistence. Berlyne (1956), in an extensive discussion of the literature, has demonstrated that these two stimulus conditions are important determinants of arousal and curiosity. In a series of experiments, he has shown that there is a direct relationship between the level of arousal and degree of curiosity subjects display, and the amount of interest and complexity of the stimuli. Clinical and experimental evidence (Hunter, 1960; Winder, 1960) suggests that increases in interest and complexity might be experienced as stressful by schizophrenics, however, and have a negative effect upon persistence.

The present study was designed to investigate the effects of three levels of psychotic disturbance, two types of motivational conditions, and three levels of task interest and complexity on the persistence of schizophrenic subjects. The following hypotheses were to be tested:

I. Persistence in schizophrenia is an inverse function of the degree of psychotic disturbance. Specifically it is hypothesized that ambulatory schizophrenics in treatment on an outpatient basis would show greater persistence on assigned tasks than chronic, hospitalized patients, and that persistence is related to the level of adjustment of patients within an outpatient sample.

II. Persistence in schizophrenics is a function of the nature of the assigned task. It is hypothesized that schizophrenics are less persistent as tasks become more interesting and complex.

III. Persistence is a function of the subjects' involvement in the task. Specifically it is hypothesized that schizophrenics would persist longer in ego-involving, than in task-oriented conditions.

METHOD

Subjects

Ninety male schizophrenic patients under treatment in Veterans Administration facilities served as subjects. They were selected without regard to the specific type of schizophrenia. The subject population was comprised of three samples of 30 patients each: two outpatient samples under treatment at the Veterans Administration Outpatient Clinic, Brooklyn, New York, and a hospital sample drawn from the Veterans Administration Hospital, Northport, Long Island. The groups were differentiated as follows: Group A—ambulatory schizophrenic outpatients who were rated on a scale described below as comparatively well adjusted; Group B—ambulatory schizophrenics who were rated on the same scale as poorly adjusted; Group C—chronic and regressed hospitalized patients, all of whom had at least 2 years of continuous hospitalization at the time of the study.

Patient Ratings

The degree of psychotic disturbance of the ambulatory patients was assessed by means of an adjustment rating scale. This scale is currently being used in an ongoing evaluation of a milieu-therapy program at the Outpatient Clinic. The scale assesses the patient's behavior in terms of such variables as motivation, interpersonal relations, self-concept, affective control, dependency, mood, use of community facilities, and family adjustment.

Interobserver reliability correlation coefficients average .80, corrected for attenuation, and the scale has successfully discriminated between patients with high and low degrees of psychotic disturbance in other research (Blumenthal, 1964a). The subject's were rated by three staff members who knew the patients at least 6 months. The 60 patients comprising the outpatient group were selected from a larger pool of papatients attending the Day Treatment Center. Patients with rating-scale scores of 3.8 or less were considered as having a low adjustment, while those with scores of 4.8 or better were considered as having a relatively high adjustment. The average of the entire patient population from which the two samples were drawn is 4.3.

Hospitalized patients were assessed for their degree of psychotic disturbance by means of the Montrose Adjustment Rating Scale (Rackow, Napoli, Klebenoff, & Schillinger, 1953). This scale has been used in a number of studies (Blumenthal, 1964(b); Hunter, 1960; Schooler & Spohn, 1960) and is a reliable method of distinguishing between regressed hospital patients and partially remitted hospital patients. The 30 patients in the hospital group received scores of 10 or less which previous research has established as providing an unambiguous cutoff point for regressed patients.

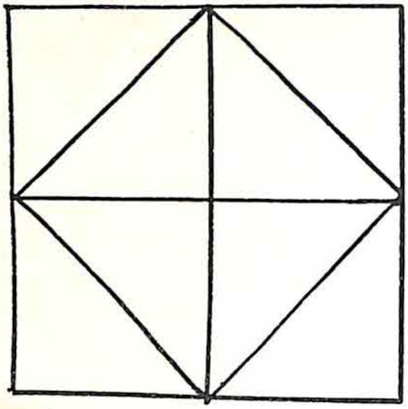


FIG. 1. Insoluble geometric design for problem-solving task.

Tasks

The three tasks used in the study were as follows: (*a*) Simple-repetitive: Filling in circles with a pencil. This is a highly repetitive and monotonous task. The subjects were presented with a stack of 8½ × 10-inch pages filled with 18 double-spaced typewritten rows of the letter "O," with 39 letters per line. The task consisted in filling the space in the Os completely with a pencil. (*b*) Digit-letter matching: Subjects were given a stack of papers as above with each sheet consisting of 18 rows of randomly ordered digits from 0 through 9. Letters and matching digits were printed on the top of the page. The task consisted in writing the appropriate letter beneath the digit in the body of the page as in the familiar Digit-Symbol procedures on intelligence tests. (*c*) Problem-solving: Conditions were similar to the preceeding tasks, with the exception that the subjects were requested to copy a geometric design.

This was to be done without lifting the pencil from the paper and without going over any line a second time. Spaces for the attempts were demarcated on each page. This task is impossible to solve.

The three tasks were ranked as to their level of interest and complexity by 5 psychologists. All of them ranked the task of filling in "Os" lowest and the problem-solving task highest in interest and complexity with the digit-letter task between them.

Motivational Conditions

General instructions. On all three tasks the subjects were instructed to work as long as they pleased and to indicate when they wanted to stop. If at any time they asked if they should continue, the same general instructions were repeated. The subjects were stopped only when they definitely expressed the desire to stop.

One-half of the subjects received ego-involved instructions and the remaining received task-oriented instructions.

Ego-involved instructions. Under this condition subjects were told that they were being tested to find out how well they could perform in comparison to other patients.

Task-oriented instructions. The task was defined as part of a study of the comparative quality of mechanical and wooden pencils under varied writing conditions, and their opinions about the pencils were solicited in a posttest inquiry. In actuality all subjects were given wooden pencils to "test." As in the ego-involved task, all subjects, as determined by questioning after the experimental session, found the explanation to be plausible and participated in the experiment seriously.

Subjects worked in a well lit, large office at a plain wooden table, with only the pile of papers before them. The experimenter sat off to the side and to their rear. Performance was timed to the nearest second and the time score was the basic unit of the analyses.

Experimental design. Each subject in the three patient groups was randomly assigned to one of the combinations of three tasks and two instructions.

TABLE 1

AVERAGE PERSISTENCE TIME IN MINUTES FOR 3 SCHIZOPHRENIC GROUPS ON THREE TASKS UNDER TWO DIFFERENT MOTIVATIONAL CONDITIONS

Level of adjustment	Tasks			*M*	
	Circles	Digit-letter	Problem-solving		
High-adjusted outpatients					
Task	27.15	21.13	11.15	19.81	22.18
Ego	37.15	24.10	12.41	24.55	
Low-adjusted outpatients					
Task	17.08	11.49	23.58	17.38	21.62
Ego	25.45	28.57	23.52	25.85	
Hospitalized patients					
Task	12.47	4.92	4.93	7.44	7.95
Ego	11.58	10.67	3.12	8.46	
M	21.81	16.81	13.13		

Note.—Task-oriented $M = 14.88$, Ego-involved $M = 19.62$.

This provided data for a $3 \times 2 \times 2$ factorial design with five different subjects in each of the 18 cells. The data were treated by analysis of variance after applying a logarithmic transformation necessitated by lack of homogeneity of variance and the skewed distribution of the time scores.

RESULTS

Group means of three diagnostic groups under the two motivational conditions on the three tasks are shown in Table 1 and the results of the statistical analysis are presented in Table 2.

Variability decreased progressively from the high-adjusted group ($\sigma = 22.8$) through the low-adjusted group ($\sigma = 16.9$) to the hospitalized group ($\sigma = 11.13$). The variability for the hospitalized subjects was significantly lower ($p < .05$) than either of the outpatient samples.

In support of Hypothesis I concerning level of adjustment there was a highly significant difference ($p < .001$) in persistence among the three groups. The difference was principally between the hospitalized patients and both outpatient samples. Although the high-adjustment group persisted slightly longer than the low-adjustment group, the difference was not significant. The combined outpatient samples averaged 21.87 minutes on all tasks in contrast to the 7.95 minutes of the hospitalized group.

As can be seen in Table 1, the average persistence time for all diagnostic groups combined decreased progressively from the least interesting and complex task through the intermediate task to the highly complex task. Despite these apparent differences in the predicted direction, the finding was not statistically significant because of the marked intragroup variability. The hypothesis of differences in persistence of schizophrenics as a function of task interest and complexity was not supported.

The average persistence scores of the ego-involved group were longer than the task-oriented group. The difference, in the predicted direction, falls at the $p = .08$ level of significance. It is believed that larger samples would clearly establish the reliability of this finding, and that the result can be tentatively considered as supporting the hypotheses.

TABLE 2

ANALYSIS OF VARIANCE OF PERSISTENCE SCORES

Source	*df*	*MS*	*F*
Tasks (A)	2	1.01	1.12
Diagnosis (B)	2	15.56	17.28*
Instructions (C)	1	2.69	2.99**
A × B	4	.90	1.00
A × C	2	1.16	1.11
B × C	2	.25	
A × B × C	4	.34	
Error	72	.90	

* $p < .001$.
** $p = .08$.

None of the interactions between tasks, motivational conditions, and diagnosis were significant. The performance of the outpatient and inpatient samples was remarkably similar across tasks except for the consistently higher scores of the outpatient group.

DISCUSSION

The results of this study indicate that ambulatory schizophrenics are clearly superior to inpatients in persisting at a variety of tasks at several levels of interest and complexity. A number of reasons, singly or in combination, may account for this finding. Hospitalized patients may become more anxious than outpatients when confronted with tasks that demand sustained effort. Increased tension may cause the more rapid onset of fatigue and consequent task interruption. Then again, interruption may represent for them a way of leaving an anxiety-laden field. Another possibility would be that hospitalized schizophrenics do not have the energy level or ego strength to focalize efforts sufficiently to persist at tasks. They may be too easily distracted by intrapsychic events to persist on externally imposed goal-directed tasks. Equally plausible is the notion that hospitalization per se may have further affected the patients' motivation for sustained effort.

Several alternative explanations may account for the finding that schizophrenics do not persist differentially with increased levels of task interest and complexity. On the basis of previous research it would be expected that normals would persist longer as tasks become more interesting, complex, and cognitively challenging. The schizophrenics, as predicted, performed exactly in the opposite direction, even though tasks did not differentiate significantly within the psychotic group as a whole. It is possible that schizophrenics, in contrast with normals, are not as responsive to differences in interest and complexity, and that these are not particularly relevant aspects of tasks for them. There is no question that the experimental tasks became progressively complex, but the two variables of complexity and interest may be confounded. Although solving problems was judged by five psychologists to be more interesting than filling in Os, there is no assurance that schizophrenics would agree with that judgment. Even more important is the recognition that the problem-solving task, being impossible to complete, might well have been experienced as a failure each time it was attempted. Future work should attempt to construct tasks in which failure experiences are controlled.

The results suggest that more attention be paid to persistence both in assessing the level of adjustment of ambulatory schizophrenics, and in judging their readiness for discharge from the hospital.

REFERENCES

BERLYNE, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1956.

BLUMENTHAL, R. L. Sociometric choice patterns in outpatient schizophrenics. *Newsletter for Research in Psychology,* 1964, **6**, 41–43. (a)

BLUMENTHAL, R. L. The effect of level of mental health, premorbid history and interpersonal stress upon the speech disruption of chronic schizophrenics. *Journal of Nervous & Mental Disease,* 1964, **139**, 313–323. (b)

EYSENCK, H. J. *The structure of human personality.* London: Methuen, 1953.

GUILFORD, J. P. *Personality.* New York: McGraw-Hill, 1959.

HUNTER, M. The effects of interpersonal interaction upon the task performance of chronic schizophrenics. Unpublished doctoral dissertation, Columbia University, 1960.

IVERSON, M. A., & REUDER, M. E. Ego involvement as an experimental variable. *Psychological Reports,* 1956, **2**, 147–181.

LEWIN, K. *Dynamic theory of personality.* New York: McGraw-Hill, 1935.

RACKOW, L., NAPOLI, P., KLEBANOFF, S., & SCHILLINGER, A. A group method for the rapid screening of chronic psychiatric patients. *American Journal of Psychiatry,* 1953, **109**, 561–566.

RYANS, D. G. The measurement of persistence: An historical review. *Psychological Bulletin,* 1939, **36**, 715–739.

SCHOOLER, C., & SPOHN, H. The susceptibility of chronic schizophrenics to social influence in the formation of perceptual judgments. *Journal of Abnormal and Social Psychology,* 1960, **61**, 348–354.

WINDER, C. L. Some psychological studies of schizophrenics. In D. D. Jackson (Ed.), *The etiology of schizophrenia.* New York: Basic Books, 1960. Pp. 191–247.

(Received May 21, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 251–261

ROTE LEARNING IN SCHIZOPHRENIC AND NORMAL SUBJECTS UNDER POSITIVE AND NEGATIVE REINFORCEMENT CONDITIONS

JANET TAYLOR SPENCE[1]

Veterans Administration Hospital, Iowa City, Iowa

LEONARD D. GOODSTEIN[2]

University of Iowa

AND CHARLES V. LAIR[3]

Veterans Administration Hospital, Knoxville, Iowa

The effects on rote learning performance of adding positive or negative verbal reinforcers to the information provided by presentation of the correct response was studied in normals and in remitted and nonremitted schizophrenics. On a verbal discrimination task, the reinforcers facilitated the performance of all groups ($p < .05$). On paired-associate lists, the reinforcers facilitated the performance of the poor performers among the normals and remitted schizophrenics and retarded it for better performers ($p < .05$). For nonremitted schizophrenics, the reinforcers produced better performance in the better performers but had no significant effects on poor performers. The varying roles played by the reinforcers as related to task and subject variables were discussed.

In an attempt to test hypotheses concerning the deviant motivational reactions of schizophrenics to social censure and reward, a number of investigators have compared the effects of different classes of verbal response-contingent reinforcers on the performance of laboratory tasks in schizophrenic and normal groups (cf Silverman, 1963). In many of these studies, the positive or negative reinforcers supply to the subject his only information about the correctness or adequacy of his response. Although demonstrations of the differential effects on performance of the positive and negative reinforcement conditions or of an interaction betwen reinforcement condition and diagnosis are typically given a motivational interpretation, the possibility that the results are due instead to differences in the kind or amount of information extracted from the reinforcers by various groups cannot be ignored.

For example, in recent studies by the present writers (Spence, 1964; Spence & Lair, 1965), the effects of two verbal reinforcement combinations on the performance of a two-alternative verbal discrimination task were investigated. In the positive or "reward" condition, the experimenter said "Right" after each of the subject's correct choices but nothing after incorrect ones (right-nothing or Rn combination) while in the negative or "punishment" condition, the experimenter said "Wrong" after incorrect choices but nothing after correct ones (wrong-nothing or Wn combination). Under conditions in which the reinforcement procedures were not explained in preliminary instructions, thus requiring that subjects discover for themselves the information obtainable from the experimenter's response or failure to respond, performance was found to be better under the Wn than under the Rn condition. The difference between conditions was larger for schizophrenics, particularly those in poor remission, than for normal subjects. Superficially, these results would seem to support the often-made generalization that schizophrenics are more motivated to avoid punishment and less motivated to obtain rewards than are normals (e.g., Atkinson & Robinson, 1961; Cohen & Cohen, 1960;

---

[1] Now at the University of Texas and Austin State School.

[2] Now at the University of Cincinnati.

[3] Now at David P. Wohl Memorial Mental Hygiene Clinic, St. Louis University.

Rodnick & Garmezy, 1957). Probability analyses, however, suggested that in neither group were the results due to the differential effects of "Right" versus "Wrong" per se. Rather they were due to the inability of the subjects, particularly the schizophrenics, to extract the appropriate information from nothing in the Rn combination as effectively as in the Wn combination. Studies of verbal conditioning in schizophrenia that have employed conceptual types of tasks have obtained the same type of interaction between diagnosis and positive versus negative reinforcement (Cohen & Cohen, 1960; Hartman, 1955; Leventhal, 1959). These results may have been due to a similar information discrepancy between conditions since the reinforcement procedures utilized in these studies are essentially of the same Rn or Wn type.

A procedure which minimizes this type of confounding of the informational and noninformational properties of social reinforcers of the right-wrong variety was employed in a study of schizophrenics and nonschizophrenics conducted by Atkinson and Robinson (1961). Using paired-associate lists, these investigators followed the usual procedure of supplying the correct response automatically after presentation of each stimulus term. This procedure is one that forces the subject to determine for himself whether his attempt to anticipate had been correct or incorrect. Thus, since no explicit evaluation of the subject's response is given, the procedure may be designated as a neutral one. In their experimental groups, Atkinson and Robinson supplemented this neutral information with social reinforcers which specifically called attention to the correctness or incorrectness of the subject's response. Schizophrenics whose failure to respond correctly were followed by such comments from the experimenter as "Wrong" or "No" were found to perform significantly better than the subjects in the nonreinforced control condition but no significant differences were found between the latter and the subjects reinforced for correct anticipations by "Good," "Right," etc. The schizophrenics' motivation to avoid punishment, it was concluded, was greater than their motivation to obtain reward. In the normal groups, on the other hand, performance was better under positive reinforcement than under negative. Unfortunately, no normal subjects were tested under the purely neutral condition so that it was impossible to determine whether the addition of both or of only one type of social reinforcement affected performance and in what direction.

In the present investigation, the effects on performance of adding verbal reinforcers to the neutral information provided by presentation of the correct response was also studied in groups of schizophrenic and normal subjects. In Experiment I the same two-alternative verbal discrimination lists (except for mode of feedback) as had been used in previous studies by the writers (Spence, Lair, & Goodstein, 1963) were employed. While the immediate purpose of the experiment was to determine if the Atkinson-Robinson results could be duplicated with a different type of rote learning task, the study is one of a series of investigations being conducted by the present writers in which task is being held constant and mode of introducing the reinforcers varied in an attempt to determine the different roles that reinforcers may play in affecting performance.

The results of Experiment I were dissimilar to those obtained by Atkinson and Robinson in several respects. In order to determine whether task factors were responsible for these discrepancies, a second experiment was conducted which replicated the Atkinson-Robinson investigation more exactly by employing paired-associate lists.

EXPERIMENT I

Method

Subjects.[4] Seventy-two long-term schizophrenic patients, all male, were selected from the Veterans Administration Hospital, Knoxville, Iowa according to the following criteria: eighth-grade education or more, 50 years of age or less, no history of alcoholism, drug addiction or CNS damage, and a minimum of 36 months of continuous hospitalization. In addition, subjects were selected according to current mental status, the symptoms of half of them having been judged to be in full or partial remission at the time of testing and of the other half in nonremission. This classification was made on the basis of ratings obtained from a nurse and nursing assistant on the subject's ward, only those subjects about whom both raters agreed being selected.

---

[4] In both Experiments I and II it was originally planned to test a total of 60 subjects in diagnostic groups not to be further subdivided (e.g., medical patients) and 90 subjects in groups that were to be divided (e.g., schizophrenics, remitted and nonremitted). Practical considerations, however, prevented reaching these target numbers in all instances.

Sixty male patients from medical and surgical wards of the Veterans Administration Hospital, Iowa City, none of whom had a history of psychiatric disturbance, served as the normal group. Except for length of hospitalization, these patients were selected by the same general criteria as the schizophrenics. The mean ages for the schizophrenic and normal groups, respectively, were 36.18 and 37.88 years, and mean education 10.80 and 10.58 years. Mean duration of current hospitalization for the schizophrenics was 72.6 and 84.1 months for the remitted and nonremitted subjects, respectively. Each of the groups was subdivided into three experimental subgroups, matched on age and education as well as on performance on a practice list. Within each schizophrenic subgroup, half of the subjects were remitted and half nonremitted, age and education being equated for the two types of patients.

Materials and apparatus. Two verbal discrimination lists, described in detail elsewhere (Spence, Lair, & Goodstein, 1963) were employed. Briefly, the first or practice list consisted of 6 pairs of familiar 4-letter words and the second or experimental list of 16 pairs of familiar 5-letter words. The lists were projected on a screen by means of a Dunning Animatic projector, using a film strip in which the presentation of each pair of words was followed by presentation of the correct word. The materials were exposed at a 4:4-second rate, with 4 seconds between trials. The strip contained 8 repetitions of the first list and 16 of the second, the pairs appearing in 4 different serial orders across repetitions with the right-left position of the members of each pair being counterbalanced across orders.

Procedure. Preliminary instructions stated that the subject was to call out one of the words each time a pair was presented, his task being to learn to select the "correct" member. List 1 was then given for 8 trials under neutral conditions (i.e., the correct word appeared on the film strip, unaccompanied by any verbalizations from the experimenter), Trial 1 being a guessing trial. List 2 was given for 16 trials, with the experimental conditions being introduced without comment or explanation on Trial 2. The experimenter said "Good," "Fine," "Right," or "Uhhuh" following each correct choice for members of the Positive groups and "No," "Not right," "Wrong" or "Uh-uh" following each incorrect choice for members of the Negative groups. The order of the experimenter's responses in each condition was determined by a prearranged random schedule. The experimenter's verbalizations were spoken in a neutral but firm tone of voice and were timed to coincide with the appearance of the correct word on the screen. For subjects in the Neutral (control) condition, the experimenter remained silent during the administration of the list, thus continuing the procedure used with all subjects on the practice list.

Results

The means for the groups of total number of correct responses per subject on Trials 2–16 of the experimental list are shown in Table 1. A 2 × 3 analysis of variance of these data comparing the schizophrenic groups (remitted and nonremitted combined) with the normal subjects under the three conditions yielded no significant terms. Since both reinforcement conditions tended to produce better performance than neutral, the data from the positive and negative conditions were combined and a 2 × 2 analysis (schizophrenic-normal, reinforcement-neutral) performed. In this analysis, the reinforcers were shown to improve performance significantly ($F = 4.37$, $df\ 1/128$, $p < .05$). The *F*s for Groups and the Groups × Conditions interaction, however, continued to be nonsignificant.

A separate comparison of the two schizophrenic groups under the three conditions yielded a significant term for conditions ($F = 7.49$, $df = 1/66$, $p < .01$), again suggesting that the presence of the reinforcers facilitated performance. As may be seen from Table 1, however, the facilitating effects of the negative condition were limited to the nonremitted subjects, the remitted subjects

TABLE 1

MEANS OF TOTAL NUMBER CORRECT RESPONSES PER SUBJECT ON VERBAL DISCRIMINATION LIST: EXPERIMENT I

Condition	Groups					
	Nonremitted subjects		Remitted subjects		Normals	
	M	*SD*	*M*	*SD*	*M*	*SD*
Neutral	188.7	48.66	214.7	25.57	213.7	21.77
Positive	217.6	35.66	224.9	8.53	218.7	18.24
Negative	213.1	18.38	212.3	28.48	217.5	22.31

performing at approximately the same level under this condition as the parallel neutral group. The interaction between groups and conditions in the 2 × 3 analysis was not, however, significant ($F = 1.61$).

The results reported above differ from those obtained by Atkinson and Robinson (1961) in several important respects. The latter investigators found an interaction between group and reinforcement condition, their normal subjects performing better with the positive reinforcers than with the negative and their schizophrenic subjects the opposite. In the present study, on the other hand, there was a suggestion of a performance difference between the two reinforcement conditions in only one of the three groups (remitted schizophrenics) and this was in the opposite direction to that found by Atkinson and Robinson.

To determine whether tasks factors were responsible for these discrepant findings, a second experiment was conducted in which, as in the Atkinson-Robinson study, paired-associate lists were employed. Since these investigators had given subjects two paired-associate lists under the experimental conditions, finding that the effects of the reinforcers on performance were greater in magnitude on the second of the two lists, this same procedure was followed in Experiment II. In part because Atkinson and Robinson had used nonhospitalized individuals as their normal subjects, it was also decided to test several comparison groups in addition to medical patients, namely college students and hospital employees.

EXPERIMENT II

Method

Subjects. A second group of 72 male schizophrenics, half remitted and half nonremitted, were selected from the Knoxville Veterans Administration Hospital by the same criteria as in Experiment I. Several nonpsychiatric groups were also tested. One consisted of 60 male patients from medical and surgical wards of the Iowa City Veterans Administration Hospital, and a second of 45 male hospital employees from the Knoxville Veterans Administration Hospital. Neither of these groups differed significantly from the schizophrenic subjects in mean age and education. The third group of normal subjects consisted of 90 students, half of them male and half female, from introductory psychology classes at the University of Iowa.

An equal number of subjects from each of the groups was assigned to the three experimental conditions (positive, negative, neutral) in such a manner that the parallel subgroups were matched on practice list (List A) performance and in the nonstudent groups, on age and education. In the student sample, an equal number of males and females was assigned to each condition, and in the schizophrenic sample, an equal number of remitted and nonremitted patients.

Materials and apparatus. Three paired-associate lists were prepared. Each consisted of 10 pairs of familiar words, as determined by a frequency count of 50 or more per million in the Thorndike-Lorge (1944) list, with a minimal degree of formal and meaningful similarity among the words within and among lists. The lists were presented by means of a Dunning Animatic projector on a film strip which contained 11 repetitions of the first (practice) list and 16 repetitions of the other two, the pairs on each list appearing in three different serial orders. The lists were given at a 3:3-second exposure rate, with 3 seconds between trials.

Procedure. After being given usual instructions for paired-associate learning, the subjects were given 10 anticipation trials on the practice list (List A) under neutral conditions, followed by 15 anticipation trials on each of the two experimental lists. For the latter (Lists 1 and 2) the two types of reinforcement conditions were instituted on the first anticipation trial in the same manner as in Experiment I. As in the Atkinson-Robinson study, both overt errors and response omissions were followed by a response from the experimenter in the negative groups.

Results

A number of subjects, particularly those in the normal groups, were responding perfectly on the two experimental lists well before the fifteenth anticipation trial. In order to minimize ceiling effects, the data from only the first 10 anticipation trials were analyzed statistically for all groups on each list.

List 1

Normal groups. A 3 × 4 analysis of variance of total number of correct responses per subject on the first experimental list revealed significant differences in performance among the normal groups ($p < .001$), the medical patients being the poorest, followed by the hospital employees, student males, and student females. The F for conditions was not significant. The pattern of the results, however,

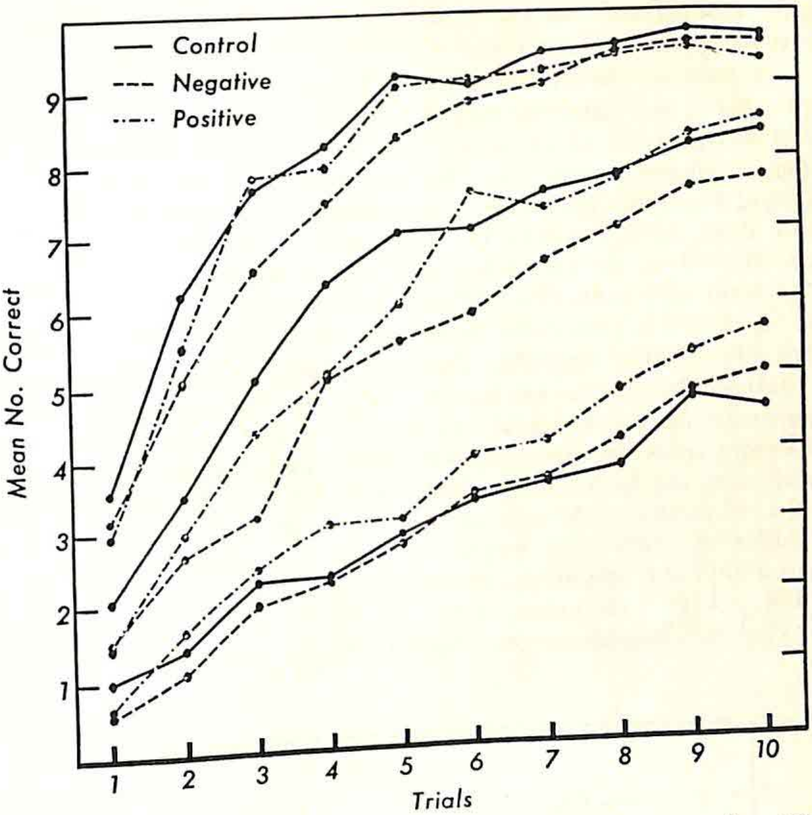


FIG. 1. Learning curves for the normal subjects at the three performance levels under the three reinforcement conditions. (First experimental list, Experiment II.)

suggested that the reinforced subjects did perform differently from those in the neutral condition but that the direction of the difference was not the same for all groups, the reinforcers tending to produce poorer performance in the college students and somewhat better performance in the medical patients. This interaction, however, did not prove to be statistically significant ($F = 1.84$).

Further inspection of the data suggested that the observed interaction between group and condition may have stemmed from the relation between the subject's group membership and his general performance level, and that the effects of the reinforcers on performance might be demonstrated more clearly if subjects were stratified according to overall performance level without regard to the subject population from which they were originally selected. The data were therefore reanalyzed in the following manner.

The performance data from the practice list (List A) were pooled for all 195 normal subjects and the total group then divided into upper, middle, and lower performance levels on the basis of subjects' total number of correct responses on that list. Since within each original sample, subgroups had been equated for List A performance, this method of division resulted in an approximately equal number of subjects from each of the condition subgroups being assigned to each of the three levels of List A performance.

The List 1 performance of the subjects, divided by condition and level, is plotted in Figure 1. It may seen that at the two upper levels, performance was deterred by the addition of positive or of negative reinforcers to the neutral information provided by presentation of the correct response, the interference effect being particularly marked in the early trials. For the poor (low level) performers, on the other hand, the reinforcers had the opposite effect, tending to facilitate performance, particularly in the later trials.

Since the magnitude of the effects of the reinforcers appeared to be related to stage of practice, two 3 × 3 analyses of variance of these data were performed, based on total number of correct responses per subject on Trials 1–5 and Trials 6–10, respectively. In the analysis of Trials 1–5, the term for conditions was significant ($F = 3.86$, $df = 2/186$, $p < .05$) but the Levels × Conditions interaction was not ($F = 2.41$), suggesting that the mild facilitating effect of the reinforcers on the performance of the low-level performers in the early trials was not sufficient to off-set statistically the interference effect produced by the reinforcers at the two upper levels. The analysis of Trials 6–10, however, did result in a significant interaction ($F = 3.22$, $df = 4/186$, $p < .05$), the nature of the interaction precluding a significant term for conditions ($F = 1.93$). In both analyses, the F for levels was, of course, highly significant ($p < .001$), as it was in all subsequent analyses to be reported.

It will also be observed from Figure 1 that at each of the three performance levels, performance was better under positive than under negative reinforcement. A 2 × 3 analysis of variance of correct responses on all 10 trials comparing the two reinforcement groups at the three levels indicated that the superiority of the positive condition over the negative was statistically significant ($F = 3.42$, $df = 1/124$, $p < .05$). There was no indication that the magnitude of the difference between the two reinforcement conditions was related to subjects' performance level, the Conditions × Levels term being nonsignificant ($F < 1.00$).

Schizophrenics. Examination of the data from the schizophrenic groups on the first

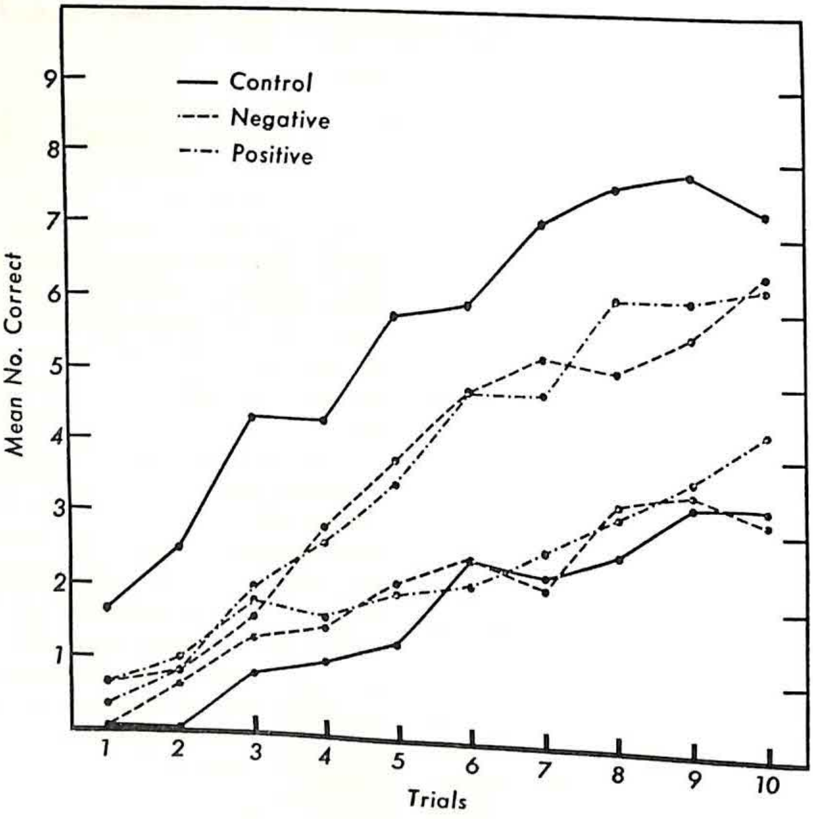


FIG. 2. Learning curves for the remitted schizophrenics at the upper and lower performance levels under the three reinforcement conditions. (First experimental list, Experiment II.)

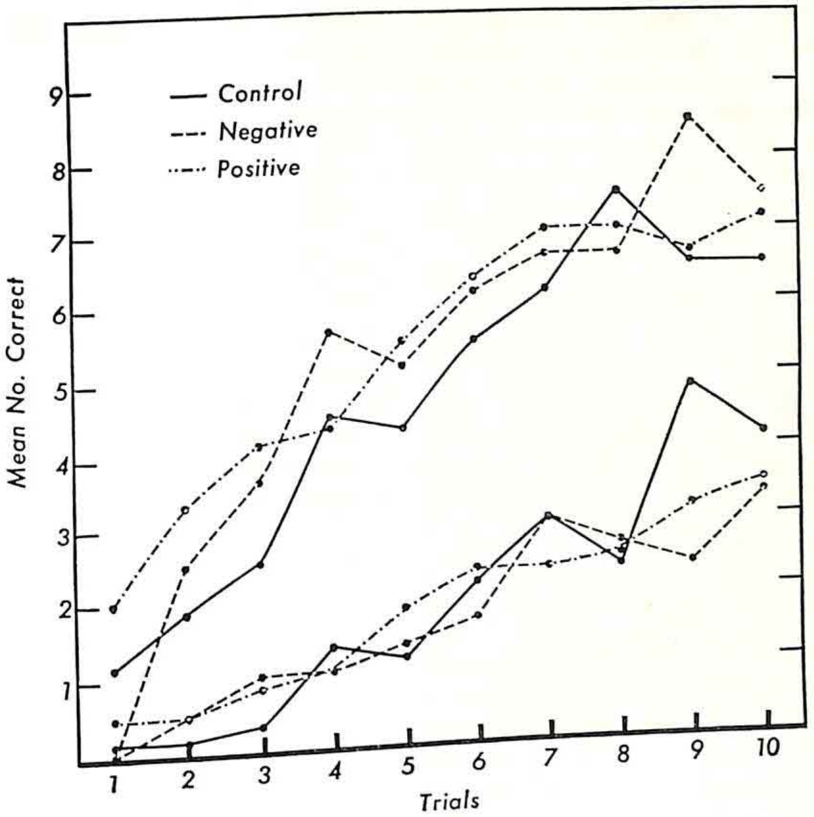


FIG. 3. Learning curves for the nonremitted schizophrenics at the upper and lower performance levels under the three reinforcement conditions. (First experimental list, Experiment II.)

experimental list suggested that it would also be useful to differentiate the subjects on the basis of performance level, as well as to analyze separately the data from the remitted and nonremitted subjects. Because of the small *N*s involved, each separate subgroup was divided into upper and lower halves on the basis of practice list performance, thus resulting in six subjects per cell.

The performance of the remitted groups on List 1, shown in Figure 2, was similar to that of the normal subjects in that the reinforcers tended to facilitate performance for the subjects at the lower level but to retard it for the subjects at the upper level. It will be further noted that, unlike what had been found with the normal subjects, performance was similar under the positive and negative conditions at each level. The data for the two reinforcement conditions were therefore combined and a 2 × 2 analysis of variance (2 levels, reinforced versus neutral conditions) of total number of correct responses per subject performed. In this analysis, the interaction between levels and conditions was found to be significant ($F = 6.17$, $df = 1/35$, $p < .05$).

Turning to the data from the nonremitted schizophrenics, it will be seen from Figure 3 that as a whole, the upper groups performed about as well as the remitted subjects. Contrary to what had been found with the latter, however, as well as with the better performers among the normal subjects, the upper level reinforcement groups were not inferior in performance to the neutral group but instead tended to be superior. Further, the reinforcement conditions seem to have had no consistent effects on the performance of the lower groups. Since there was no indication of a difference between positive and negative conditions, the two were combined and a

2×2 analysis of variance of total number of correct responses per subject performed. Neither the F for conditions nor for Conditions × Levels was significant.

List 2

The effects of the reinforcers on performance of the second experimental list were not as marked as on List 1. For the normal subjects, however, the pattern of the results was the same, the addition of the reinforcers to neutral information interfering with the performance of the subjects at the upper levels and facilitating it for subjects at the lowest level. As on List 1, the Levels × Conditions interaction for Trials 6–10 was significant ($p < .05$). A comparison of the positive and negative groups also indicated that at all three levels, performance continued to be better under the former condition ($.05 < p < .10$).

Analyses of the data from the remitted and from the nonremitted schizophrenics showed that whether the reinforcement conditions were combined and compared with the neutral condition or whether all three conditions were considered separately, neither the F for conditions nor the Conditions × Levels interaction was significant for either group. With respect to the trends of means, however, the performance of the upper halves of each group was similar to what was found on List 1: the reinforcers facilitated performance for the better performers among the nonremitted subjects and retarded it for the remitted. A 2×2 analysis of variance comparing remitted and nonremitted subjects at the upper level under reinforced and neutral conditions indicated that this interaction between condition and type of patient was significant ($p < .05$). For the lower level groups, there were no consistent performance differences among conditions in either the remitted or nonremitted subjects.

DISCUSSION

In many studies comparing the effects of classes of verbal reinforcers on the behavior of schizophrenic and normal subjects, the assumption is made that any differences in performance between conditions or diagnostic groups are due to the differential capacities of the reinforcers to motivate the subjects or to arouse them affectively. As several writers (Farber, 1955; Lazarus, Deese, & Osler, 1952) have pointed out, however, both positive and negative reinforcers may have a number of properties instead of, or in addition to, motivational and affective ones, depending on such variables as the specific nature of the reinforcers themselves, the manner in which they are introduced into the experimental situation, task variables, and type of subject. Thus it cannot be automatically assumed that all differences in performance between conditions or groups are attributable to motivational variables. For example, the writers cited earlier the results of several reinforcement combination studies (Spence, 1964; Spence & Lair, 1965) which suggested that performance differences that were obtained under positive and negative conditions and the interaction between conditions and the subjects' diagnostic status were due to informational rather than to motivational factors.

The writers would like to suggest that the results of the present investigation might also be more adequately accounted for in nonmotivational terms. In developing an alternate type of explanation for the effects of the reinforcers on performance, the data obtained from the normal subjects in Experiment II will first be considered.

For the normal subjects, it will be recalled, the addition of social reinforcers to the information provided by presentation of the response term tended to deter the performance of better performers but to facilitate it for the poorer performers. The reinforcers, it will be assumed, played two contradictory roles during the acquisition of the paired-associate lists. The first was to act as a distractor, in the same manner as an irrelevant, extraneous stimulus which diverts the subject's attention from the task and thus interferes with efficient performance. The second role of the reinforcers, it is hypothesized, was to emphasize and call attention to the information available from the mechanical presentation of each response term, particularly the relation between that term and the correctness or incorrectness of the subject's attempt to anticipate it. Utilization of this kind of supplementary information had the effect of facilitating performance.

The appearance of net facilitation or interference in performance was related to the subjects' performance level. The better performers, who by definition were performing effectively, appear to have found the reinforcers unnecessarily redundant as sources of information so that for these subjects, the reinforcers acted primarily as distractors. On the other hand, poor performers, who are likely to have difficulties in concentrating and in comparing the correct response with their attempts to anticipate it, were able to benefit from the additional information being provided so that although they may also have found the reinforcers somewhat distracting, their overall performance showed facilitation effects. Thus the major difference between the poor and better performers in the experimental conditions may have been the degree to which they required the reinforcers to assist them in paying close attention to the task and the details of their performance. A similar suggestion concerning the beneficial role that may be played by reinforcers in subjects suffering from attentional difficulties has been made by Losen (1961).

Turning to the data from the schizophrenic groups, the performance of the remitted patients on the first experimental list was similar in pattern to that of the normal subjects: for the better performers, performance under the reinforcement conditions was inferior to that under the neutral condition while for the poorer performers, performance under the reinforcement conditions was superior. For the latter group, however, this facilitation effect disappeared on the second experimental list, performance being similar under all three conditions.

While the results obtained from the normal subjects and remitted patients were generally similar, the nonremitted schizophrenics exhibited a different pattern. For these subjects, the better performers exhibited performance facilitation rather than retardation under the reinforcement conditions on both paired-associate lists while the poorer performers among the nonremitted subjects did not respond differentially to the various conditions.

It will be noted that the groups for which the reinforcers had only a transient, if any, influence on performance were the poor performers among the two schizophrenic groups. It may be argued that these subjects, who were observed to be particularly erratic in their trial-to-trial performance, were the sickest patients, too preoccupied and distracted by internal processes to give to the task the close concentration and attention its mastery required or even to profit from the supplementary information being provided by the reinforcers. Following this line of reasoning further, the better performers among the nonremitted schizophrenics may have suffered from more attentional difficulties than the better performers among the normal subjects and remitted patients, yet be intact enough to take advantage of the reinforcers and thus at least partially overcome these defects. Hence, unlike the other groups, the better performers among the nonremitted patients showed performance facilitation rather than retardation under the experimental conditions.

On the verbal discrimination list employed in Experiment I, it will be recalled, the reinforcers resulted in performance facilitation for each of the groups, schizophrenic and normal. These data were reanalyzed to take account of the subjects' performance level by dividing each subgroup into the upper and lower halves on the basis of practice list performance. In each diagnostic group it was found that the facilitation effect noted in the undivided sample came largely from the poor performers, the discrepancy in performance between the neutral and reinforced subjects on the experimental list being particularly marked among the nonremitted schizophrenics. These were the subjects, it will be remembered, for whom the reinforcers were found to have no significant effects in Experiment II. The trends of the results among the better performers in each group were also in favor of the reinforced subjects, thus providing no evidence that for these subjects, the presence of the reinforcers inhibited performance, as they did on the paired-associate lists.[5]

---

[5] The most convincing evidence for this type of interaction with task for the better performers came

Several differences between the two types of tasks seem to be important in accounting for the differences in the two sets of results. The verbal discrimination task was both simpler and more highly structured than the paired-associate task in that it required *recognition* of the correct member of each pair of words rather than *recall* of the specific response associated with a given stimulus from among a number of alternatives. Mastery of the verbal discrimination list thus demanded less active effort and less sustained concentration, and may therefore have made the subjects' performance less vulnerable to any distracting effects the reinforcers may have had. The same task factors may have been responsible for the capacity of the poorer performers among the schizophrenic groups, especially the nonremitted patients, to benefit from the presence of the reinforcers in the verbal discrimination situation, when they could not do so with the paired-associate lists. These patients required the supplementary assistance that could be provided by the reinforcers more than other subjects, but only in the more highly structured, less demanding verbal discrimination situation were they able to take advantage of them.

As is implied in the discussion above, the addition of positive and negative reinforcers to neutral information quite consistently affected performance in the same direction within a given subgroup of subjects. Comparing the effects of the two reinforcement conditions directly, the performance of the normal subjects under the positive condition was found to be superior to their performance under the negative condition on the paired-associate lists of Experiment II. Thus the results lend some confirmation to the earlier findings of Atkinson and Robinson (1961). While these differences could be given a motivational interpretation, it should be recalled that on the verbal discrimination list of Experiment I, no evidence of a similar difference was obtained. It is possible that in a two-alternative situation, calling attention to an incorrect response is equally effective as calling attention to a correct one. On the other hand, in anticipation learning, with its multitude of response alternatives, emphasizing the subject's correct responses may provide more useful information than noting his errors, especially since the bulk of these errors are response omissions.

For the schizophrenic groups, there were no significant differences between positive and negative conditions in either Experiment I or II, although the trends that occurred also tended to favor the former. Thus on neither the verbal discrimination nor the paired-associate lists did the evidence confirm the findings obtained by Atkinson and Robinson which showed that schizophrenics performed better under negative than under positive reinforcement. Subtle differences between the two sets of investigations with respect to such variables as subject characteristics, the experimenter's tone of voice in delivering reinforcers, details of list construction, length of anticipation interval, etc., could all be responsible for these discrepancies.

REFERENCES

Atkinson, Rita L., & Robinson, Nancy M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, **62**, 322–326.

Cohen, E., & Cohen, B. D. Verbal reinforcement in schizophrenia. *Journal of Abnormal and Social Psychology*, 1960, **60**, 443–446.

Farber, I. E. The role of motivation in verbal learning and performance. *Psychological Bulletin*, 1955, **52**, 311–327.

Hartman, C. H. Verbal behavior of schizophrenics and normal subjects as a function of types of social reinforcement. Unpublished doctoral dissertation, University of Iowa, 1955.

Lazarus, R. S., Deese, J., & Osler, Sonia F. The effects of psychological stress upon performance. *Psychological Bulletin*, 1952, **49**, 293–317.

Leventhal, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *Journal of Abnormal and Social Psychology*, 1959, **59**, 162–166.

Losen, S. The differential effects of censure on the problem-solving behavior of schizophrenic and normal subjects. *Journal of Personality*, 1961, **29**, 258–272.

---

from a direct comparison of the groups of remitted schizophrenics and of medical patients tested in Experiments I and II. For the better performers among both types of patients, performance under the reinforcement conditions was found to be inferior on the paired-associate lists to that under the neutral condition but somewhat better on the verbal discrimination list.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

SILVERMAN, J. Psychological deficit reduction in schizophrenia through response-contingent noxious reinforcement. *Psychological Reports,* 1963, **13** (Monograph Supplement 2-V13), 187–210.

SPENCE, JANET T. Verbal discrimination performance under different verbal reinforcement combinations. *Journal of Experimental Psychology,* 1964, **67**, 195–197.

SPENCE, JANET T., & LAIR, C. V. The effect of different verbal reinforcement combinations on the verbal discrimination performance of schizophrenics. *Journal of Personality and Social Psychology,* 1965, **1**, 245–249.

SPENCE, JANET T., LAIR, C. V., & GOODSTEIN, L. D. Effects of different feedback conditions on verbal discrimination learning in schizophrenic and nonpsychiatric subjects. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 339–345.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received May 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 262–265

ALCOHOLISM AND PSYCHOLOGICAL DIFFERENTIATION:

EFFECT OF ALCOHOL ON FIELD DEPENDENCE[1]

STEPHEN A. KARP,[2] HERMAN A. WITKIN, AND DONALD R. GOODENOUGH[3]

Department of Psychiatry, Downstate Medical Center, State University of New York

This is the 1st of a series of studies of stability of perceptual field dependence (taken as an indicator of level of differentiation) during the alcoholic cycle. These studies are in preparation for a longitudinal study of the development of alcoholism. In this 1st study, alcohol ingestion failed to have a significant effect on performance in 2 tests of field dependence, the rod-and-frame and body-adjustment tests. In a 3rd test, the embedded-figures test, speed of locating a simple figure in a complex design was reduced under the influence of alcohol, but this change seems attributable to reduced ability to concentrate under alcohol rather than to a change in perception per se.

Recent studies have shown that alcoholics, both men and women, are markedly field dependent in their perception and tend to have a global concept of their bodies (Bailey, Hustmyer, & Kristofferson, 1961; Karp, Poster, & Goodman, 1963; Witkin, Karp, & Goodenough, 1959). Both characteristics have been conceived as manifestations of limitedly developed differentiation, and they have been found to occur in association with such other manifestations of limited differentiation as a poorly developed sense of separate identity and a tendency to use nonspecific defenses as primitive denial and massive repression (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962). The clustering of these characteristics, or their opposites, suggests that relatively limited or relatively developed differentiation is a pervasive feature of individual psychological functioning.

Their field-dependent perception and their global concept of the body suggest that alcoholics tend to be limitedly differentiated. This relationship is, however, amenable to two kinds of interpretation. The first, which we have labeled the "predisposing hypothesis," suggests that limited differentiation was a characteristic of alcoholics prior to their development of that symptom and may have, in some way, contributed to their later alcoholism. The second, which we have called the "consequence hypothesis," would suggest that the limited differentiation found to characterize most alcoholics is, in some way, a consequence of their drinking behavior (e.g., a result of the physical deterioration that is brought about through the combination of excessive drinking of alcohol and insufficient diet).

From the evidence now available we cannot tell whether level of differentiation is a predisposing factor or a consequence of alcoholism. There is, however, considerable evidence from studies of nonalcoholic subjects that measures reflecting relative level of differentiation are highly stable over time. One kind of evidence of stability has come from a longitudinal study in the period of 10 to 24 years of age.[4] Another kind has come from studies which sought, without success, to alter performance on tests of field dependence by various sorts of experimental intervention, including exciting and depressing drugs, with which alcohol is sometimes compared.

Were we justified in extending these findings of stability to alcoholics, they would provide considerable support for the "predisposition hypothesis." Such results would imply that the low level of differentiation, which has been found to characterize alcoholics studied after they had been drinking for many years, should also have characterized them before their drinking began. In the absence of such justification, definitive evidence on whether limited differentiation is predisposing to alcoholism or follows from that condition can only come from a "prediction study" in which an assessment of differentiation is made in early adolescence, prior to the onset of alcoholism, the incidence of alcoholics among those who were more differentiated and less differentiated is later determined, and level of differentiation among those who subsequently become alcoholics is assessed. If limited differentiation is in fact predisposing toward alcoholism, we would have a useful method of identifying potential alcoholics, thereby contributing to the possible prevention of alcoholism, and we would also have some further insight into the development of this condition.

---

[1] From the Alcohol Research Unit and the Psychology Laboratory. We are indebted to Carol Eagle, Beverly Goodman, and Peggy Plumeau for their assistance in administering the test battery.

This work was supported in part by Grant MH-00628-11 from the National Institutes of Health, United States Public Health Service.

[2] Now at the Sinai Hospital of Baltimore and The Johns Hopkins University School of Medicine, Baltimore, Maryland.

[3] Research Career Development Awardee (Level 2; Grant MH-K3-16,619), National Institute of Mental Health, United States Public Health Service.

[4] Unpublished study by Witkin, Goodenough, and Karp.

In preparation for a prediction study, which alone can be conclusive, there is obvious merit in pursuing relevant specific problems amenable to short-range solution. Such preparation is valuable, and indeed necessary, in longitudinal studies because they are so time consuming and irreversible. A particularly important issue to consider in advance of a prediction study of alcoholism is the stability of characteristics of differentiation during the various phases of the cycle through which the alcoholics may pass. If these characteristics show ready change in varying phases, the "consequence hypothesis" would become more plausible as an interpretation of the known difference between alcoholics and nonalcoholics and would reduce the strategic need for a prediction study.

In the present series of papers we report several studies dealing with the problem of stability. Specifically, we have explored the effects on one indicator of level of differentiation—extent of perceptual field dependence—of acute intoxication, of continuous alcoholism over many years, and of prolonged sobriety after an extended history of alcoholism. If field dependence or, more broadly, limited differentiation is a consequence of alcoholism, field dependence might be greater in the intoxicated state than in the sober state; greater among alcoholics with a long history of alcoholism than among those with a short history; and greater among alcoholics who have continued to drink than among those who have stopped drinking.

The present study deals with the effect of acute intoxication on field dependence.

METHOD

The subjects were 24 male alcoholics, recruited from a larger group of alcoholic inpatients at Kings County Hospital in Brooklyn, New York. Testing was done at least 4 days after admission to the hospital (usually following a period of heavy drinking) in order to allow some of the acute toxic effects of drinking to wear off. The subjects ranged in age from 27 to 57 years, with a mean age of 40.8 years.

Each subject received a battery of three tests of perceptual field dependence on two different occasions, once while sober (dry condition) and once after ingestion of controlled doses of alcohol (wet condition). In the wet condition, the subject was given a "cocktail" consisting of fruit juice and 1 milliliter of absolute alcohol per kilogram of body weight. A second cocktail, of the same alcoholic content, was administered approximately 2 hours after the first. Blood samples were taken at approximately half-hour intervals to determine level of alcohol content of the blood. This level varied slightly from subject to subject, and for the same subject, from one blood sample to the next. However, approximately 100 milligrams of alcohol per 100 milliliters of blood were found in most of the samples taken. With regard to behavioral manifestations of intoxication, there was considerable variety among subjects and, in some cases, within the same subject. Some showed no apparent behavioral change or at most mild slurring of words or mild euphoria, while others became nauseated, vomited, and showed great euphoria. For two subjects, the effect of the alcohol was so severe as to prevent completion of the test battery. These subjects were not included in the study.

The test-retest interval ranged from 1 to 7 days. In most cases both test and retest were conducted while the subject was an inpatient; however, some subjects were discharged prior to retesting and returned by appointment to complete the study. Half the subjects, randomly chosen, were tested first while sober and retested with alcohol, and half were tested in the reverse order. The order in which the three tests were administered was randomly determined for each session for each subject.

The battery of tests of perceptual field dependence consisted of the embedded-figures test (EFT), the rod-and-frame test (RFT), and the body-adjustment test (BAT).[5]

---

[5] Instructions for administering and scoring these tests are reported in outline by Witkin et al. (1962) and in detail by Witkin (1948, 1950).

TABLE 1

Summary of *F* Tests for Effects of Intoxication, Retesting, and Their Interaction

Source	*df*	*F* ratios		
		RFT	BAT	EFT
Alcohol (A)	1/22	1.94	0.37	9.79*
Retesting (order) (R)	1/22	1.79	1.17	29.10*
A × R (Drug sequence)	1/22	0.23	0.93	3.30

*$p < .01$.

The EFT consists of a series of 24 complex geometric designs. Embedded in each complex design is a simple figure which the subject must locate as quickly as he can. The subject's score is the total time he takes to locate the simple figures in the 24 complex designs. The longer the subject takes to find the simple figures, the more field dependent is his perception.

In the RFT, the subject is seated in a totally darkened room. The only objects he can see are a luminous rod surrounded by a luminous square frame; the rod and the frame are always presented in tilted positions (to left or right). The subject's task is to adjust the rod to the vertical while the frame remains in its initial position of tilt. On 8 trials, the subject's chair is tilted to the right, on 8 trials to the left, and on 8 trials the subject sits erect. The subject's score for the test is the sum of degrees deviation of the rod from the true upright for the 24 determinations. Larger scores reflect greater field dependence.

In the BAT, the subject is seated in a chair within a small, specially constructed room. Both room and chair can be tilted, independently or together, to left or right. At the outset of each of 6 trials, chair and room are tilted to set positions. The subject's task is to adjust the chair (and thus his body) to the vertical while the surrounding room remains tilted. The score for the test is the sum of degrees deviation of the chair from the upright on the 6 trials, larger scores reflecting greater field dependence.

In each of these three tests the task is to separate an object (a geometric figure, a rod, or the body) from an organized field. Extent of field dependence is reflected in relative ease of separating object from field.

Results and Discussion

A separate analysis of variance was carried out for each of the three perceptual tests, evaluating the effects of intoxication, retesting, and the interaction between intoxication and retesting (drug sequence). A summary of these analyses is presented in Table 1.

There is no evidence that alcohol, retesting, or drug sequence had any effect on performance in the BAT or RFT. However, in the EFT subjects required significantly more time to locate the simple figures in the complex designs after ingestion of alcohol than before. Mean time per figure was 130 seconds under alcohol as compared to 106 seconds when sober.

It seems reasonable to attribute the difference between the EFT and the other two tests to the fact that the EFT is a speed test. In order to find the simple figures in the EFT, the subject must focus his attention on the test material for considerable periods of time. Failure to concentrate or lapse of interest result in higher (more field-dependent) scores. In contrast, the RFT and BAT are not timed and do not require long periods of concentration. When tested "wet," subjects often joked, cried, and engaged in other kinds of irrelevant behavior. Whereas such behavior could obviously affect time scores on the EFT, it could not directly affect the "degrees-deviation" scores on the BAT and RFT.

The absence of significant differences in BAT and RFT scores under the wet and dry conditions makes it reasonable to conclude that alcohol ingestion does not affect perceptual field dependence. The significant change in EFT scores with alcohol, as suggested, cannot be attributed to a change in perception as such.

The only other significant effect revealed by the analysis of variance is a retesting effect for the EFT. The subjects were able to find the simple hidden figures significantly faster on retesting than on initial testing. A testing effect was not found for either the BAT or RFT. The difference between these two tests and the EFT is again attributable to a difference in test format. In the EFT, finding the simple figure identifies the correct response for the subject; he is able to use this information to find the simple figure more quickly the next time around. On the BAT and RFT, subjects are not able to tell whether the adjustments they make are correct or not, so that their experience is not likely to benefit them on subsequent testing.

The results of this study suggest that extent of field dependence in alcoholics is not different when they are intoxicated or sober. These findings are consistent with those of other studies which have shown that level of field dependence is not changed by drugs (Franks, 1956; Pollack, Kahn, Karp, & Fink, 1960), by special training (Gruen, 1955; Witkin, 1948), by stress (Kraidman, 1959), or by hypnosis.[6]

The finding that extent of field dependence is not altered by alcohol ingestion does not, as noted, provide direct support for the view that the marked field dependence associated with alcoholism antedates the onset of drinking behavior. However, by reducing the plausibility of one alternative hypothesis (that field dependence is a consequence of alcoholism) to account for the difference between alcoholics and nonalcoholics, it increases the plausibility of the predisposing hypothesis and hence the value of a more definitive prediction study.

---

[6] Unpublished study by Jerome Schneck and Rochelle Wexler.

REFERENCES

BAILEY, W., HUSTMYER, F. E., & KRISTOFFERSON, A. Alcoholism, brain damage and perceptual dependence. *Quarterly Journal of Studies on Alcohol,* 1961, **22**, 387–393.

FRANKS, C. M. Differences déterminées par le personalité dans la perception visuelle de la verticalité. *Revue de Psychologie Appliquée,* 1956, **6**, 235–246.

GRUEN, A. The relation of dancing experience and personality to perception. *Psychological Monographs,* 1955, **69** (Whole No. 399).

KARP, S. A., POSTER, DOROTHY C., & GOODMAN, A. Differentiation in alcoholic women. *Journal of Personality,* 1963, **31**, 386–393.

KRAIDMAN, EMMA. Developmental analysis of conceptual and perceptual functioning under stress and non-stress conditions. Unpublished doctoral dissertation, Clark University, 1959.

POLLACK, M., KAHN, R. L., KARP, E., & FINK, M. Individual differences in the perception of the upright in hospitalized psychiatric patients. Paper read at Eastern Psychological Association, New York, April 1960.

WITKIN, H. A. The effect of training and of structural aids on performance in three tests of space orientation. Civil Aeronautics Administration, Division of Research, Technical Report, 1948, No. 80.

WITKIN, H. A. Individual differences in ease of perception of embedded figures. *Journal of Personality,* 1950, **19**, 1–15.

WITKIN, H. A., DYK, RUTH B., FATERSON, HANNA F., GOODENOUGH, D. R., & KARP, S. A. *Psychological differentiation.* New York: Wiley, 1962.

WITKIN, H. A., KARP, S. A., & GOODENOUGH, D. R. Dependence in alcoholics. *Quarterly Journal of Studies on Alcohol,* 1959, **20**, 493–504.

(Received June 18, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 266–269

HYPNOTIC AGE REGRESSION:

A TEST OF THE FUNCTIONAL ABLATION HYPOTHESIS [1]

JAMES R. LEONARD [2]

University of Kentucky

The hypothesis, that responses acquired after the time to which S is hypnotically age regressed are functionally ablated, was tested by utilizing a modification of the retroactive interference paradigm with a paired associates task. 5 preselected female Ss learned List A (S_1-R_1). 2 weeks later they learned List B (S_1-R_2) whereupon they were hypnotically age regressed to the date of original learning, when R_1 strength was high, and given further S_1-R_1 trials. Appropriate A-A and A-B-A comparison groups ($N = 10$ each) were employed. The finding that the hypnotic age-regression procedures failed to preclude the occurrence of interlist intrusions in the relearning of S_1-R_1 does not support the hypothesis.

When a hypnotized adult is given the suggestion that he is a child, his ensuing behavior, including his performance on psychological tests, often appears appropriate for the suggested age. The nature of this hypnotic age regression has been a matter of controversy since the phenomenon was first reported by Krafft-Ebing (1888). The phenomenon has been recently subjected to critical review by Gebhard (1961) and by Barber (1962).

The present study was designed to test the functional ablation hypothesis which holds that when a subject is hypnotically age regressed, events of his past are experienced as for the first time and responses acquired after the time to which he is regressed are temporarily unavailable.

Edmonston (1961) noted that previous tests of the hypothesis proved inconclusive since either no objective measures had been obtained at the time to which subjects were later regressed or, when such measures were available, they could readily be simulated by overcooperative subjects. He correctly indicated that a study was needed which would utilize prior measures in an involuntary response context. While Edmonston undertook such a study and concluded that his results supported the functional ablation hypothesis, the definitiveness of this conclusion has been questioned by Barber (1962) and Leonard (1963).

The present experiment employed a modification of the retroactive interference paradigm with appropriate A–A and A–B–A comparison groups. A 2-week age regression period was used; the presence or absence of retroactive interference with previously established learning serving as the test of the functional ablation hypothesis. It follows from this hypothesis that in pair associates tasks involving a high degree of interlist response competition, the subjects who have learned S_1-R_1 and later learned S_1-R_2 should show no interference from the latter when hypnotically age regressed to the time when S_1-R_1 strength was high.

METHOD

Subjects

Twenty-five females, employees of the Veterans Administration Hospital or wives of University of Kentucky students, served as the subjects. They ranged in age from 21 to 30 years and all had at least the equivalent of 2 years of college. Five of these subjects were selected from the available pool of volunteers as having demonstrated clinical evidence of hypnotic age regression to childhood. These five subjects comprised the experimental group. Twenty subjects selected from the same pool formed matched A-A and A-B-A comparison groups, paired on the basis of initial learning performance. Volunteers rejected from the experimental group were not permitted to serve in the comparison groups.

The criteria for choosing the experimental subjects and the procedure used to induce hypnotic age regression were quite intricate. As they followed closely those employed by Edmonston (1961) and are reported elsewhere (Leonard, 1963), they will not be repeated here. It should, however, be mentioned that while all of the phenomena described by Wolberg (1948) were attempted with each of these subjects, the criteria of deep trance were: (*a*) the ability to open the eyes and still maintain the trance, (*b*) posthypnotic amnesia and the performance of posthypnotic suggestions. Trance induction for all subjects was by the visual fixation technique until the posthypnotic suggestion for deep trance entry in response to a signal from the experimenter was successfully carried out.

---

[1] This article is adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Kentucky. The author expresses his appreciation to Frank A. Pattie, Chairman of the dissertation committee, and to the staff of the Veterans Administration Hospital, Lexington, Kentucky.

[2] Now at Bedford Veterans Administration Hospital, Bedford, Massachusetts.

Prior to the experiment proper, the subjects of the comparison groups were seen for two half-hour sessions designed to acquaint them with the subject role. During these sessions they recalled events of their childhood and performed simple addition and subtraction tasks. All subjects were led to believe that they were participating in a series of unrelated experiments.

Apparatus

Two lists of 11 nonsense syllables of low-association value (Glaze, 1928) were paired with a series of two digit numbers. The numbers served as stimulus items and were the same for both lists: (S_1-R_1), (S_1-R_2). To minimize position effects four orders of each list were constructed. These orders were typed on continuous white tapes and used with a Gerbrands memory drum. Learning was by the paired-associates method with 3-seconds exposure time and an intertrial interval of 9 seconds.

Procedure

Date 1. After the subject had learned List A (S_1-R_1) to the criterion of one errorless trial she was questioned regarding her earlier activities of the day. A Polaroid snapshot was then taken which the subject dated and kept as a memento of her participation. The taking of this snapshot was assumed to constitute a memorable event in the experience of the subject and would serve as a reference point to which the experimental subjects would later be hypnotically regressed. The subject was then told to close and rest her eyes and that afterward she would continue to learn the list. Ten seconds later the subject was told to open her eyes and get ready. It was emphasized that rather than say nothing she could call out the syllable that came to mind even if she felt it might be wrong. Further practice was then given until the subject again reached one errorless trial.

Date 2. Two weeks later the subject returned to the same room at which time the memory drum was out of sight. The subject was handed a pencil and paper and asked to write all the syllables that she could recall from the previous session. A syllable was counted correct if spelled or pronounced as on Date 1. Two minutes were allowed for this reproduction. Thereafter, the groups were treated differentially as follows:

Experimental group. The subjects of this group were given nine trials on List B (S_1-R_2), after which the drum was put aside as the experimenter commented, "Well that's all for that." Using a pre-established rapid induction technique, the subject was placed in a deep trance and regressed to the date of original learning of List A. When this date was reached the subject was stopped and questioned about her activities preceding the learning session. This questioning continued until 15 minutes had elapsed since the beginning of trance induction.

The subject was then told that she had just learned a list of number-syllable pairs, had her picture taken, and would soon open her eyes and continue to learn the list. At this point it was again emphasized that rather than say nothing she should call out any syllable that came to mind, even if she felt that it might be wrong. The subject was then told to open her eyes and get ready. Thirty seconds later the memory drum was turned on and (S_1-R_1) trials were continued until the criterion of one errorless trial was reached. When the criterion had been reached the subject was again questioned as to the date. All subjects gave the day and date on which the original learning took place. The subject was then brought out of the trance and questioned at length regarding her understanding of the experiment.

A-B-A group. The subjects of this group were also given nine presentations of List B. They were then handed blank covered booklets containing pictures from the Meier Art Judgment Test and given the following instructions:

> We are trying to see if we can determine which factors are most important in a person's liking a picture. There are some pictures in these booklets. Each page has two pictures which are slightly different. I want you to indicate which one pleases you most. When you finish one booklet, I'll give you another.

The subject spent 15 minutes looking at the pictures. She was then stopped and told that further practice would be given on the list that she had learned 2 weeks ago (S_1-R_1). At this point it was again emphasized that, rather than say nothing, she should call out any syllable that came to mind even if she felt that it might be wrong. Practice was continued until one errorless trial had been attained.

A-A group. The subjects of this group received no practice on List B. Instead, they spent an equivalent amount of time viewing the Meier Art Judgment pictures. These subjects spent a total of 25 minutes viewing the pictures, after which they were treated in the same manner as the subjects of the A-B-A group.

RESULTS

Statistical checks were made to determine whether the groups differed in initial-learning ability as indicated by the amount of error and number of trials to criteria on Date 1 (see Table 1). In no instance was the mean

TABLE 1

TRIALS AND ERRORS TO CRITERIA IN LEARNING LIST A ON DATE 1 AND LIST A RESPONSES CORRECTLY REPRODUCED ON DATE 2

Group	N	Prephoto Learning				Postphoto Learning				Reproduction Date 2	
		Trials		Errors		Trials		Errors		Items Correct	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A–A	10	18.9	3.31	89	18.05	3.8	1.67	6.1	2.08	4.8	1.03
A–B–A	10	18.2	2.74	86.3	16.58	3.9	1.79	5.3	2.58	4.7	1.06
Experimental	5	18.4	.89	91.2	8.40	4.4	1.52	5.8	1.92	5.0	.71

difference between any of the groups sufficient to yield statistical significance. Hence, we may assume that the groups were equated in learning ability.

All subjects denied having rehearsed the items during the 2-week interval and, as the mean differences between the groups in the number of List A responses correctly reproduced on Date 2 failed to reach statistical significance, we may assume that the groups did not differ in retention (see Table 1).

The relative performance of the groups in relearning List A on Date 2 is the major factor of interest in the present study. The following between group comparisons were made: (*a*) number of trials and errors to criterion, (*b*) number of errors made on the first relearning trial, (*c*) cumulative error on the first four relearning trials, (*d*) type of errors made.

The trial and error data for relearning on Date 2 appear in Table 2. Comparison of the mean differences by t test revealed that the A-A group made significantly less error than both the A-B-A ($p < .02$) and experimental ($p < .01$) groups. The A-A group also required significantly less trials to reach criterion than both the A-B-A and experimental groups ($p < .01$). By neither measure did the A-B-A and experimental groups differ.

Table 2 also presents the amount of error made by each group on the first and first four relearning trials. Between-groups comparisons by t tests revealed that the A-A group made significantly less error than both the A-B-A group ($p < .01$) and the experimental group ($p < .01$) on the first trial. The A-A group also made significantly less error than both the A-B-A group ($p < .05$) and the experimental group ($p < .01$) during the first four trials. By neither measure did the performance of the A-B-A and experimental groups differ.

The frequency of interlist intrusions occurring in the relearning performance of the A-B-A and experimental groups was examined and it was found that in this respect the groups did not differ (see Table 3). It should be emphasized that each subject of the experimental group made at least one interlist intrusion error.

Qualitative results. The experimental subjects had been told, while still hypnotized at the end of relearning on Date 2, that they

TABLE 2

DATE 2 RELEARNING OF LIST A: ERRORS ON TRIAL 1, ERRORS ON TRIALS 1–4, TOTAL ERRORS AND TRIALS TO CRITERION

Group	N	Errors Trial 1		Errors Trials 1–4		Total errors		Total trials	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A–A	10	7.6	1.38	19.3	3.61	22.4	9.86	5.8	1.13
A–B–A	10	9.6	1.06	23.6	5.23	30.7	9.33	8.3	1.89
Experimental	5	10.0	.63	26.0	3.16	34.2	6.10	8.6	1.14

TABLE 3

COMPARISON OF INTERLIST INTRUSIONS MADE BY THE A–B–A AND EXPERIMENTAL GROUPS WHILE RELEARNING LIST A ON DATE 2

Group	*N*	Range	*M*	*df*	*t* Value
A–B–A	10	3–6	4.7	13	.85
Experimental	5	2–7	5.2		

would have full posthypnotic recall for the events that had transpired during the trance. Then, in the normal waking state, they were carefully questioned and their responses noted. First they were asked what they thought the experiment was about. Their comments indicated that they all believed that they had been participating in a series of unrelated experiments. Hints regarding the nature of the experiment were given to the extent of reminding them that they had relearned List A in the trance. Even after these hints, only three of the subjects mentioned some relationship between the hypnotic procedures and the learning tasks. They "guessed" that the experiment was a study of relative learning ability in the trance and waking states. All three expressed the belief that hypnosis would enhance ability.

DISCUSSION

As the functional ablation hypothesis states that responses acquired after the time to which the subject is hypnotically regressed are temporarily unavailable, the finding of interlist intrusions in the relearning performance of all subjects of the experimental group constitutes a serious challenge to the hypothesis. Further, since the A-B-A and experimental groups did not differ in the number of interlist intrusions or efficiency of relearning, we may state that the age-regression procedure neither prevented nor reduced interference from learning acquired after the time to which the experimental subjects were hypnotically regressed. We are, therefore, led to conclude that the functional ablation hypothesis is not supported.

Platonow (1933) believed that behavior elicited during hypnotic age regression represented an organic reproduction of engrams formed at an earlier age. He viewed hypnotic age regression as a powerful tool for the investigation of memory. More recently, Reiff and Scheerer (1959) have relied heavily upon data obtained during hypnotic age regression in developing their model of the memory process. Yet the utility of the hypnotic procedures, if not the worth of the data obtained with their use, remains open to question; for it has still to be adequately demonstrated that a hypnotically regressed subject can perform behaviors "that could not be performed voluntarily by an appropriately motivated but unhypnotized adult [Barber, 1961, p. 191]."

REFERENCES

BARBER, T. X. Experimental evidence for a theory of hypnotic behavior: II. Experimental controls in hypnotic age-regression. *International Journal of Clinical and Experimental Hypnosis,* 1961, **9**, 181–193.

BARBER, T. X. Hypnotic age-regression: A critical review. *Psychosomatic Medicine,* 1962, **24**, 286–299.

EDMONSTON, W. E. An experimental investigation of hypnotic age-regression. *American Journal of Clinical Hypnosis,* 1961, **3**, 127–138.

GEBHARD, J. W. Hypnotic age regression: A review. *American Journal of Clinical Hypnosis,* 1961, **3**, 139–168.

GLAZE, J. A. The association value of nonsense syllables. *Journal of Genetic Psychology,* 1928, **35**, 255–267.

KRAFFT-EBING, R. V. *An experimental study in the domain of hypnotism.* Stuttgart: Enke, 1888. Translated from the German by C. C. Chaddock. New York: Putnam, 1889.

LEONARD, J. R. An investigation of hypnotic age regression. Unpublished doctoral dissertation, University of Kentucky, 1963.

PLATONOW, K. I. On the objective proof of the experimental personality age regression. *Journal of General Psychology,* 1933, **9**, 190–209.

REIFF, R., & SCHEERER, M. *Memory and Hypnotic Age Regression.* New York: International Universities Press, 1959.

WOLBERG, L. R. *Medical hypnosis.* Vol. 1. New York: Grune & Stratton, 1948.

(Received June 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 270–280

LOOKING, RECALLING, AND GSR AS A FUNCTION OF DEFENSE[1]

LESTER LUBORSKY, BARTON BLINDER
School of Medicine, University of Pennsylvania
AND JEAN SCHIMEK
Downstate Medical Center, State University of New York

10 pictures were viewed for 10 seconds each, during which the eye fixations were photographed and GSR was recorded. An independent assessment was made for each of 16 Ss of the defenses of isolation and repression. We discovered that the defense of isolation was strikingly associated with venturing to look around more, and the defense of repression with looking around less: in contrast to repressors, isolators looked at the "ground" as well as the figure; they scattered their fixations and they made bigger jumps from 1 fixation to the next. Isolation was better than repression for the recall of sexual content. No strong trends emerged from correlations of GSR and defense ratings. Some looking measures correlated significantly with standard Rorschach scores. The findings seem consistent with results for cognitive styles as well as with clinical understanding of the defenses.

Not much is known about whether defenses entail altered sense organ reception of *external* stimulation, although there is much consensus that defenses involve changed awareness of *internal* stimuli, i.e., thoughts and feelings. A person whose main defense is repression might limit his *awareness* of a stimulus, but not necessarily his *sensing* of a stimulus—awareness of the memory might become attenuated *after* the taking in of a threatening stimulus and the tripping off of a danger signal. A similar view is held about people whose typical defense is isolation, except that for them the thoughts about a threatening stimulus are allowed into awareness, while the affects are submerged. These are, at least, the usual views of two common defenses. Even the large literature on perceptual defense leaves the impression that the sensing of a stimulus is not the locus of defense, but rather the form of presentation into awareness. If we could get adequate measures of defenses and measures of sensory operations, we could look into this question.

In the past decade a promising line of research in perception has revealed "cognitive styles" determining consistencies in cognitive behaviors (Gardner, Holzman, Klein, Linton, & Spence, 1959). This work has begun to indicate that the intake of stimulation differs for people with different defenses (Holzman, 1962). People who were inclined to use the cognitive style termed "leveling" were the ones who, most typically, used the defense of repression (Holzman & Gardner, 1959). (Leveling refers to the person's tendency to diminish, i.e., to "level" stimulus differentiation, as opposed to "sharpening" stimulus differentiation.) Another cognitive style, "scanning," was found to be associated with the defense of isolation—people who were extreme scanners were found to be extreme isolators (Gardner et al., 1959, p. 135). The concept of scanning was not derived from eye-movement behavior, as the term might suggest, but from performance on tests such as size estimation. It refers to deployment of attention to many aspects of a stimulus field and a preference for differentiated experiencing of the world.

A sample of the same subjects as the

---

[1] This investigation was supported in whole by Grant M-3654 from the National Institute of Mental Health, United States Public Health Service. Riley Gardner of the Menninger Foundation lent us a sample of Rorschachs on which we tried out the reliability of the defense ratings. Norman Mackworth of Harvard University helped in the conducting of the original experiment and operated the Mackworth eye camera. Howard Shevrin, John Paul Brady, and Martin Orne read a version of the manuscript and offered suggestions for further analyses. Donald Phoenix made a large contribution to the data analysis. Freda Greene assisted in many ways, but especially editorially and in preparing the recall protocols.

Gardner et al. (1959) experiment was included in a tachistoscopic study (Luborsky & Shevrin, 1962). Repressors as compared with isolators behaved differently in their memory for a tachistoscopic flash of a neutral picture. Those who *lost* information after a day's delay had been independently classed as repressors; those who *increased* information after a day's delay were classed as isolators.[2]

These findings on cognitive styles and forgetting as a function of defenses might be a product of special ways of *looking* at the stimulus. However, so far in all these studies utilizing defenses, the exact location of the person's eye fixations has not been scrutinized. We therefore focused on the question of how fixation behavior differs for people with certain typical defenses.[3] After a recent experiment (Luborsky, Blinder, & Mackworth, 1963) on the relationship of looking and recalling to increased autonomic response, as shown via GSR, we saw we had data which could resolve this question. All we needed was to add an independent estimate of each person's defensive tendencies.

PROCEDURE

Since details can be found in the earlier article (Luborsky, Blinder, & Mackworth, 1963), we give here only a brief review: 10 pictures were each exposed for 10 seconds, during which the eye fixations were photographed. The Galvanic Skin Response (GSR) was continously recorded during the exposure of the picture and during the first recall of the 10 pictures (Recall 1). The subject then chose 2 pictures he liked and 2 he disliked, and then ranked 5 pictures that created anxiety. A day later, without warning, he was asked to recall the pictures again (Recall 2).

The 10 pictures (each 7¼ × 9¼ inches) were on a holder 28 inches in front of the subject. These 10 pictures were classified by the experimenters before the experiment, on the basis of their content, into three groups: three neutral, three sexual, and four aggressive pictures.

A Mackworth eye camera photographed both the picture and the point-of-regard simultaneously. A small spot of light reflected from the cornea of the eye indicates the point-of-regard. This marker spot reflects back into the camera and is recorded on the film. The angle of deflection of the spot on the cornea alters as the subject explores the stimulus. The spot has no distracting effect upon the inspection of the picture; the picture itself, as the subject looks at it, is in no way affected by the eye photography. The apparatus has been described by Mackworth (1960) and Mackworth and Kaplan (1963). The version of the camera we used was the one clamped in the fixed position, rather than head-mounted. The accuracy of the aiming was ±1 degree of expected location of the spot. Our scored record included not only where the fixations were on the picture, but their duration and sequence.[4]

Records of 16 subjects were used. The group of 5 males and 11 females included students, housewives, and secretaries. The average age was 25. (Data of 4 other subjects had to be discarded because of temporary operational defects in the movie camera.) Standardized instructions (see the earlier article) consisted essentially of a free inspection set to merely "look at each picture." The subjects were not told the purpose of the experiment and most were unaware that their fixations were being photographed.

Looking (Eye Fixation) Measures

Inspection of "ground." For each picture we decided which were the main figural elements and designated the rest as nonfigural elements or "ground." We incorporated these decisions in a set of templates for scoring all the pictures. The measure of ground is therefore the number of frames during which the person looked at the ground, i.e., the part of the 10-second period spent on ground.

Dispersion from center. We made up a bull's-eye, dividing the picture into a center area, a middle ring, an outer ring, and the remainder, and then counted the number of frames of film (i.e., $\frac{1}{16}$-sec. units) during which there were fixations in each sector. The concentric circles had radii equal to 20% of the horizontal width of the stimulus. The center circle was drawn from the geometric center. The only crude final score used was the sum of the time spent on the two central sectors divided by the total

---

[2] An analysis of variance showed that the association of repressor and isolator with increaser and decreaser was significant ($p < .02$, Fisher's exact test). Also, decreasers tended to be levelers, and increasers tended to be sharpeners ($p < .02$, Fisher's exact test).

[3] The relationships of cognitive styles and looking behavior will be reported separately in a collaborative study with Samuel Messick and Harold Schiffman of Educational Testing Service.

[4] One aspect of the procedure may influence the results in a way that is not immediately apparent. The first card seen by each subject as part of the necessary operation to check the alignment of the apparatus is a fixation card containing a white cross. When the experimental cards are shown, the subject is told that now he is free to look at what he likes. The fact of having been asked on the fixation card to look at the center of the cross may predispose some subjects to continue to look there, regardless of the succeeding liberating instruction. Subjects with different basic defense types may have unequal capacities to liberate themselves to look where they like. However, against this factor is Buswell's work (1935), in which there was no initial fixation point required, showing that the typical subject looked first in the center.

TABLE 1

Intercorrelation Matrix of Looking (Eye Fixation) Measures ($N=16$)

Measure	Ground score	Dispersion from center	Scatter	Scatter per fixation	Track length	Duration per fixation
"Ground" score						
Dispersion from center	.00					
Scatter	.33	.43				
Scatter per fixation	.37	.63**	.79**			
Track length	.31	.67**	.64**	.89**		
Duration per fixation	.27	.35	−.25	.10	.32	

** $p<.01$.

viewing time. Actual values of this score were from 0 to 1, i.e., degrees of centrality; however, in all correlations the variable was reflected to indicate dispersion.

Scatter (cells covered). A grid was constructed dividing up the picture into 144 squares. Each cell was approximately 0.006 of the total area. The score was the number of cells containing one or more fixations.

Scatter per fixation. This measure was the number of cells containing one or more fixations divided by the total number of fixations on the picture. The measure is the same as the area per fixation and is therefore the inverse of density. Scatter per fixation correlates .79 with the pure scatter measure.

Track length. This is the mean length of the track (i.e., "saccade,") from one fixation to the next in the sequence as the fixations were made.

Duration per fixation (mean). This is the mean number of frames during which the person maintained each fixation. Since the subject had 10 seconds in which to watch each picture and the film moved at 16 frames per second, 160 frames passed during inspection of each picture.

All of these looking measures were selected because they might be facets of venturing to look around the picture versus not venturing to look around. An intercorrelation matrix (see Table 1) shows that actually not all of these have high correlations with each other. Dispersion, scatter, scatter per fixation, and track length are moderately to highly correlated and probably measure the tendency to focus versus scan visual fields (and may relate to the cognitive style of focusing-scanning). The ground score and the mean duration per fixation are markedly less related to the other measures.

Recall Measures

Number of pictures forgotten. This score can range from 0 to 10. Ordinarily, there were no more than two or three pictures forgotten.

Items recalled. The content of each picture was reduced to information units. The number of units varied from about 8 to 18 items. The most usual number was around 10. Almost all these items were quite specific parts or meanings of the picture, so that very little choice was left to the judge as to how the recall should be scored. Reliability of scoring on similar types of material has been adequate or better, e.g., Shevrin and Luborsky (1958). Two scores were presented: (*a*) number recalled of each type of item (sexual, aggressive, and neutral) across all pictures; (*b*) number recalled of each type of item across all pictures, but weighted. The weighting requires adding 1 credit or taking away 1 credit, depending upon whether there was obvious and strong acceptance or rejection of the drive expression. Only 18 of these weightings were carried out; therefore *a* and *b* are very similar. The specific scoring rules for weighting are: (*a*) if the subject mentions a sexual or aggressive content only for the purpose of denying it, or along with strong denial, 1 point is subtracted from his score; (*b*) if the subject mentions a particular sexual or aggressive item with special emphasis, 1 point is added; (*c*) if the subject mentions a sexual or aggressive item with great uncertainty, 1 point is subtracted.

GSR Measures

GSR peak (rank; during looking at each picture). This is our main measure of GSR responsivity; each subject's 10 pictures were ranked in terms of it. The "Peak" is the largest change in conductance during the 10 seconds of inspecting a picture. (The unit is the rank of the conductance change as measured in millimeters of excursion on the paper.)

GSR "nonspecifics" rate (rank; during recall of each picture in Recall 1). The measure represents the rate of discrete conductance changes during the recall of each picture. Since each subject's recall period varied in length, the nonspecifics rate measure was more appropriate than a single peak measure.

Defense Measures

We relied upon two ways of assessing defenses: the Gardner et al. (1959) method and the Levine and Spivack (1963, 1964) method.

The Gardner method (G) has the advantage over other methods of estimating defenses[5] of being closest to clinical definitions of these defenses. One looks for evidences of repression and isolation in the Rorschach and then rates the degree of each defense. (The judge also had available stories to three TAT cards and age, sex, and occupation.) These defenses do not appear directly in the Rorschach; however, if a person has depended upon a particular defense for many years, certain characterological and thought-organizational qualities are apt to be present. For example, a person who has utilized repression is apt to show "constriction of ideation, absence of intellectualizing, unreflectiveness, naiveté, relatively unmodulated affect." The specific ways these qualities might appear in a Rorschach are listed in Gardner et al. (1959, pp. 130–136): (*a*) expressive and phobic verbalizations, (*b*) lack of specificity, (*c*) failures, (*d*) poor integrative efforts, (*e*) childlike material, (*f*) symbolic content without awareness of its significance, (*g*) little variety of content, (*h*) fewer than 25 responses, (*i*) fewer than 2 *M*s (movement response), (*j*) emphasis on color. The defense of isolation has been given less space in the psychoanalytic literature than repression. "Isolation" is usually taken to mean the tendency to separate affect from idea, since the concomitance and interconnection of both in consciousness might create intolerable anxiety. Habitual use of isolation might show up in "excessive emphasis on operations of logic, attentiveness to spatial and temporal attributes of objects . . .; similar qualities appear in isolators' relationships with people [p. 133]." (Specific criteria listed in Gardner et al. are mainly the opposites of those listed for repression plus (*a*) rumination about symmetry, (*b*) intellectualizing, (*c*) high $F+\%$.) Repression and isolation are supposed to have opposite effects, awareness being narrowed by repression and broadened by isolation, according to Schafer (1948). Schafer's view is based upon broad experience with a battery of diagnostic psychological tests and is shared by many who rely upon these tests.

---

[5] Two other ways of assessing defenses were considered but not tried: Byrne's (1961) Repression-Sensitization Scale, and Shannon's (1962) clinical ratings of defense by the hospital staff on the basis of patient's behavior, for patients on whom there is consensus. In none of these, including Gardner et al. (1959), is one dealing with individual instances of repression and isolation, *but rather with the assumed*

The Levine and Spivack method (1964) uses a weighted sum of points for each response in the Rorschach. Points are given according to the degree to which each response is *specific, elaborated, incorporates movement*, and *contains impulse-content* or *self-references*. A specific response is opposed to a general or vague one; e.g., "animal" reflects more repression than "bear." Elaboration refers to adjectives and adverbs. All movement is taken to mean lessened repression. Reference to sexuality, hostility, anality, or dependence is taken to mean less reliance upon repression. The mean score of the weighted points for each response is therefore an inverse measure of reliance upon a repressive form of defense. (It is reflected in the present paper to make a high score mean high repression.) Levine and Spivack call this score the "Rorschach Index of Repressive Style (RIRS); we will refer to it as "Repression^{L}" to distinguish it from Gardner's "Repression^{G}." Some recent unpublished results by Levine (personal communication) suggest the value of also using a "Total Points" score, "Repression^{Lt}," which is uncorrected for length of the Rorschach, as well as Repression^{L}.

Reliability of defense measures. Gardner et al. (1959) did not provide the usual data on reliability. Two judges pooled their judgments into a single decision on each subject and then concentrated their analyses upon those subjects who, by judges' agreement, showed reliance upon repression or isolation. In our present experiment we obtained *for all* subjects (rather than only for extreme agreed-upon repressors versus isolators) a rating on a 1-to-5 scale for degree of reliance upon each defense. Before committing ourselves to this method of judging defense, we borrowed from Dr. Gardner a sample of 26 of his cases and had them rated blindly by Dr. Schimek. The cases happened to include, according to Dr. Gardner's division, 20 repressors and 6 isolators. Dr. Schimek's task was first to decide whether the person was primarily a repressor or primarily an isolator, and then to rate him on the degree of the defense. (At the time of making his independent ratings, he had no inkling that we were going to relate looking and defense behavior.) The results showed high agreement between the Schimek and Gardner et al. ratings. A 4-fold table for the Gardner et al. versus the Schimek classification into repressors versus isolators gave a chi-square of 23.8 ($p < .01$). In all, Dr. Schimek correctly classified 15 of the 20 repressors and all 6 of the isolators. The 5 mislabeled by Dr. Schimek as isolators were each given very low isolation scores—three of them "1"; and two, "2." If we consider only the 15 classified by both Dr. Schimek and Dr. Gardner et al., as repressors (and therefore we have *ratings* on them by both raters), the correlation is .87, which has a p value of $< .01$.

For the Levine and Spivack method (1964) scorer reliability is high (above .90 in several tests) and test-retest reliability after 1 week is .91. For periods up to 10 years, retest reliability coefficients average .60 for children.

Correlations among defense measures. The ratings by Dr. Levine on our sample were done recently, after the first version of the paper had been written, and then were incorporated into it—with no major alterations in the conclusions! The judge had no knowledge of the relationships between looking and defense. We had no specific expectations about how closely the Gardner versus Levine measures would correlate—the degree of similarity of these two methods, both purporting to measure repression, was hard to judge from the written accounts.

The two methods of estimating repression correlate significantly with each other: Repression^{G} correlates .54 with Repression^{L} and .66 with Repression^{Lt}. These are surprisingly high correlations, considering these are somewhat different methods of estimating repression.

Isolation correlates negatively with repression, as

---

effects of a chronic and generalized reliance on these defenses. One rarely even comes near capturing individual instances of defenses at their moments of operation, although one access route has been explored through instances of momentary forgetting during free association (Luborsky, in press).

was expected: −.58 with Repression^G, −.58 with Repression^L, and −.44 with Repression ^{Lt}. The negative correlations between repression and isolation show that these defenses are largely, though not entirely, antithetical. (The amount of either defense is not a matter of one's "pathology"—it is noteworthy that neither defense is significantly related to Dr. Schimek's ratings of degree of pathology.)

Defense measures in relation to Rorschach scores. What has been impressive about the Gardner et al. ratings has been the extent to which judges can make reliable ratings even though the precise weighting of the component Rorschach aspects is not specified. It is also not known to which aspects of the Rorschachs repression and isolation ratings actually relate. Conceivably, none of the standard Rorschach scoring categories would correlate significantly with these ratings; they may heavily be based upon style of verbalization. We decided, therefore, to correlate some of the standard Rorschach scores with the ratings of defense. According to Gardner et al. 1959, only a few of these concrete scores are among the criteria for making the ratings of defense. For Repression^G we have included the number of responses, variety of content, number of *M* responses, and ratio $FC/CF+C$. The same scores, plus high $F+$, are relevant to rating isolation.

Table 2 shows some high correlations of Rorschach scores with Repression^G, but no significant ones for Isolation^G. We may not have included Rorschach scores relevant to judging isolation; isolation may be mainly based upon "non-concrete" aspects such as doubting, obsessiveness, and other qualities of verbalization. Table 2 is especially striking since the range of Repression^G scores is not as great as those for Isolation^G; that is, we have more extreme isolators in our group than extreme repressors. The significant correlations fit into expected patterns; for example, the combination of number of small details, rare details, and space responses correlates −.74 with the Repression^G ratings, meaning that the more one is a Repressor^G, the *less* one gives these responses. The more one is a Repressor^G, the fewer good form responses are present (−.62) and the less variety of content is present (−.55). A sum of the concrete Rorschach Repression^G scores and a sum of the concrete Rorschach Isolation^G scores each correlates significantly in opposed directions with the defense measures, as expected.

All three of the repression measures have a roughly similar pattern of correlations with the Rorschach scores (see Table 2). The main exception is that Repression^L and Repression^{Lt} have larger correlation coefficients than Repression^G for scores connected with number of responses on the Rorschach. Repression^{Lt} is even more highly correlated with many Rorschach scores than Repression^L. This is to be expected for some of the correlations with the Rorschach scores, since Repression^{Lt} is uncorrected for number of responses—Repression^{Lt} correlates −.90 with number of responses on Rorschach. However, the correlation of Repression^L with number of responses remains moderately high even after the correction for number of responses (−.60).

RESULTS AND DISCUSSION

Defense in Relation to Looking, Recalling, and GSR

Expected Relationships between Variables

Based on our understanding of the nature of repression and isolation, we came to certain rather uncertain expectations about the interrelations of these variables:

1. Repressors might be more disposed to not look around a picture, while isolators might be more venturesome in their looking. Therefore, any measures which happened to correspond to looking around versus not look-

TABLE 2

CORRELATIONS OF RORSCHACH SCORES WITH MEASURES OF DEFENSE AND LOOKING ($N=16$)

	R^a	W	D	Dr Dd S	D/W	M^b	$\frac{FC}{CF+C}$	$F+^c$	$\frac{F+^c}{R}$	Variety of content	Sum[d] repression Rorschach scores	Sum[e] isolation Rorschach scores
Repression^G	.12	−.09	−.49*	−.74**	−.27	−.45	−.19	−.62**	−.35	−.55*	.63**	−.65**
Repression^L	−.60*	−.17	−.55*	−.34	−.12	−.72**	.21	−.55*	−.30	−.50*	.56*	−.54*
Repression^{Lt}	−.90**	−.49*	−.75**	−.66**	.01	−.69**	−.02	−.83**	−.32	−.67**	.78**	−.75**
Isolation^G	.34	−.08	.35	.30	.10	.24	−.11	.31	.19	.18	−.23	.26
"Ground" score	.30	−.01	.48*	.13	.06	.10	.23	.23	.06	−.08	−.18	.19
Dispersion	−.20	−.21	−.09	−.14	.13	−.15	.06	−.19	−.15	.04	.09	−.10
Scatter	.64**	.28	.66**	.30	.03	.43	−.08	.56*	.33	.23	−.45	.45
Scatter per fixation	.69**	.12	.68**	.65**	.12	.48**	−.17	.68**	.47	.36	−.58*	.63**
Track length	.56*	.12	.55*	.37	.10	.41	−.08	.54*	.39	.38	−.43	.47
Duration per fixation	.09	−.17	−.17	.09	.20	.13	−.05	−.01	.23	.07	.02	.04

* $p <.05$.
** $p <.01$.
[a] Number of responses on the Rorschach.
[b] Includes several *FM* and *m*.
[c] Includes not only good form ($F+$) but moderately good form ($F\pm$).
[d] Concrete scores: variety of content, R, M, $\frac{FC}{CF+C}$.
[e] Concrete scores: *ditto*, plus $F+\%$.

ing around should discriminate repressors and isolators. Looking mainly at the main figural elements should be more common in repressors, as well as less dispersion from the center of the picture, less scatter, shorter track-length from one fixation to another, and possibly looking longer in any one spot (that is, longer duration per fixation).

2. Repressors might recall less than isolators. This is based on the assumption that if one looks around less, one will recall less. Repressors might also have more difficulty recalling the sexual content.

Defenses with Looking Measures

Defenses with looking at "figure" versus "ground." The correlations in Table 3 bring to light the fact that people with different defense structures deploy their fixations differently in inspecting a picture. Isolation^G ratings correlate significantly, .55, with mean ground score for each subject; repression measures are nonsignificant.

Defenses with dispersion of fixations from center. Conceivably, the looking behavior we have just revealed has relatively less to do with looking or not looking at the figural components of the picture than it has to do with funneling versus dispersing one's fixations from the center of the picture. However, this possibility is not borne out—the correlations of dispersion with defense measures are nonsignificant.

Defenses with scatter (cells covered). With Repression^L and Repression^{Lt}, the correlations are large (−.70 and −.73, respectively); with Isolation^G and Repression^G the correlations are nonsignificant (.40 and −.13).

Defenses with scatter per fixation. The correlations are similar to those for scatter, except that the correlation for Repression^G with scatter per fixation almost reaches significance at the 5% level (−.45).

Defenses with track length between fixations. Correlations of defense measures with track length are very similar in size and direction to those for scatter and scatter per fixation.

Defenses with duration per fixation. None of these correlations in Table 3 approach signficance.

TABLE 3

CORRELATIONS OF DEFENSE RATINGS WITH MEASURES OF LOOKING *for Each Subject* ($N=16$)

Measure	Repression			Isolation
	R^G	R^L	R^{Lt}	I^G
"Ground" score	−.23	−.12	−.25	.55*
Dispersion from center	.09	−.32	−.28	.17
Scatter (cells covered)	−.13	−.70**	−.73**	.40
Scatter per fixation	−.45	−.62**	−.70**	.51*
Track length	−.37	−.61**	−.63**	.51*
Duration per fixation	−.19	−.14	−.04	.02

* $p \leq .05$.
** $p \leq .01$.

In summary, correlations for our looking measures show one main trend: the more one isolates, the more one ventures to look about; conversely, the more one represses, the more one does not look about. Significance is reached by Repression$^{L, Lt}$, but not by Repression^G. The measures which show this most clearly are scatter per fixation, scatter, track length, and to a somewhat lesser extent, ground score. When we examine the correlations picture by picture we find two sexual pictures: Number 4 (TAT 13MF) and Number 9, fitting this trend most for all the defense measures. Compared with the Repression^G measure, the Repression$^{L, Lt}$ measures tend to give somewhat higher correlations for those two pictures and for other pictures as well. (It may be significant that pictures Number 4 and Number 9 are both sexual and each depicts a man and a woman together, with an ambiguous sexual implication. Picture Number 7 is the other sexual illustration—possibly its greater obviousness distinguishes it from the other two.) Clearly, however, these results for repression and isolation hold not only for sexual pictures, but also for aggressive pictures occasionally, and even for the neutral ones. The trend may be stronger for threatening pictures, but is not confined to them.

Defenses with Recall Measures[6]

We had expected that repressors and isolators might differ in the total amount they

---

[6] A few correlations of looking measures with the recall measures were significant, but these are diffi-

TABLE 4

Correlations of Defense Ratings and Recall Measures ($N = 16$)

Recall measures Number recalled, each type item:	Repression			Isolation
	R^{G}	R^{L}	R^{Lt}	I^{G}
Sexual	−.42 (−.63)**	−.62** (−.69)**	−.70** (−.75)**	.58* (.70)**
Aggressive	−.25 (−.14)	−.04 (.10)	.14 (.30)	.18 (−.02)
Neutral	−.36 (−.35)	−.27 (−.27)	−.32 (−.32)	.34 (.34)
Total	−.42 (−.48)*	−.34 (−.25)	−.34 (−.32)	.43 (.40)

Note—Correlations in parentheses are based upon *weighted* item totals. See section "Defenses with Recall Measures." All are two-tailed tests. Since direction of these is in line with prior expectation, one-tail test might have been applicable, but we preferred to hold to the more conservative levels.

* $p \leq .05$.

** $p \leq .01$.

recalled and in the recall of neutral, sexual, and aggressive pictures and contents of these types. The repressors might have a harder time and the isolators an easier time recalling the sexual pictures and contents. This view fits descriptions of repression and isolation (Schafer, 1948). What emerges from correlation of repression and isolation ratings with recall scores is roughly in line with expectations (see Table 4):

1. The total amount recalled of all types of *items* is opposite in direction for repression versus isolation ratings, but does not reach significance. (Correlations in terms of number of *pictures* forgotten are not included, since so few pictures are forgotten in entirety.)

2. The sexual items alone correlate significantly positive with Isolation^{G} (.58) and significantly negative with Repression^{L} (−.62) and Repression^{Lt} (−.70), but just below the .05 significance level for Repression^{G}, −.42.

3. When a few of the sexual items are slightly weighted for extreme sexual emphasis, all the correlations with defense measures are increased. The correlation with isolation and number of sexual items recalled, for example, is raised from .58 to .70.[7]

Defenses with GSR

Our conception of repression and isolation led us to expect differing responses on GSR measures to the neutral, sexual, and aggressive pictures. The repressors might show more GSR response[8] (i.e., possibly implying "upsetness") than the isolators to the sexual pictures, and less response to the neutral pictures. The isolators might be more upset by the aggressive pictures and not so bothered by the sexual pictures.

The evidence for this in our correlations is slight—mostly nonsignificant with an occasional exception. For each type of picture (sexual, aggressive, and neutral) the correlations of repression and isolation with GSR peak are nonsignificant, with one very significant exception—for the sexual pictures Repression^{Lt} correlates .63 with GSR peak ($n = 15$). The implication of this correlation is in line with our supposition that repressors might show *more* GSR response to the sexual pictures than the isolators. The other correlations, though not reaching significance levels, are consistent in direction. Furthermore, and

---

cult to interpret. For example, the number of sexual items recalled correlates significantly with the scatter measure (.50); the number of aggressive items recalled correlates significantly negatively with duration per fixation (−.67). (Recall 2 was too poorly recorded to provide adequate measures.)

[7] In addition to the weighting described under "Recall Measures" we tried (independently by Barton Blinder) one other type, *a scaled recall score* for each of the three sexual and each of the four aggressive pictures. The method will not be described in detail, since it gave similar results to those from the weighting of instances of extreme emphasis.

[8] In the earlier article (Luborsky et al., 1963) we found differences on looking and recall measures for the two pictures giving the highest GSR response and the two pictures giving the lowest GSR response for each person. Our tentative interpretation of these differences was in terms of high GSR being responsible for avoidance behavior. Taking the differences for four of these measures and combining them by converting them to z scores yields a sum which can be correlated with our defense scores. The hypothesis would be that the bigger the z score, the greater the repression. This is exactly what emerges—Repression^{L} correlates .52 with the z score and Isolation^{G} correlates −.48 (which just misses significance). In a similar experiment using Heart Rate, avoidance behavior was also suspected (Luborsky, Blinder, & Mackworth, 1964).

also consistent, among the correlations for each individual picture the highest is $-.64$ for GSR peak with Repression^G for picture Number 6 ("two-farmers scene"), which is a neutral picture. This correlation means that the more subject is a repressor, the less GSR peak response he gets from this picture. The subjects with high repression ratings may have found this picture a relief, especially after the four anxiety-arousing preceding ones. (Their own anxiety and like-dislike ratings of each picture fit with this suggestion.) Correlations of GSR nonspecifics rate during recall of the picture and the defense measures are not significant.

Looking at TATlike Pictures versus Responding to Rorschach Cards

Some sizable correlations emerged between our looking measures and standard Rorschach scores (see Table 2). All of the significant correlations fit with expectations, although no specific predictions were made in advance. The highest correlations are with scatter per fixation, scatter, and track length—e.g., the higher the scatter per fixation, the more responses in the Rorschach (.69); the more the D (large, usual areas) (.68); the more the $dr + dd + S$ (rare areas, tiny areas, and white space) (.65); the more the M (movement response) (.48); the more the $F+$ (good form quality) (.68). The $F+\%$ just misses the 5% level (.47). The main implication is that high scatter per fixation, scatter, and track length while looking, goes with high productivity and diversity of perceptual response.

Taking four of the Rorschach scores which, according to Gardner et al. (1959), go with Isolation^G, converting each to Zs and then correlating the total with the looking measures, yields, as expected, a significant positive correlation with scatter per fixation. Doing the same for five Repression^G Rorschach scores yields a significant positive correlation with scatter per fixation. (Track length and scatter follow the same directions but do not reach significance.)

It seems justified to conclude that some of the same factors which determine these measures of looking behavior also determine some of the standard Rorschach scores. Very likely, one of the underlying factors is the person's style of looking, rather than only a particular attitude or set, since these looking measures on our 10 TATlike pictures were correlated with Rorschach scores on Rorschachs administered by another person at another time.

Illustration of Records of Two Subjects: Mrs. R. versus Miss I.

The scored eye-fixation records of Mrs. R. (Repression^G Rating 4, R^L Rank 2, R^{Lt} Rank 1; Isolation^G Rating 2) and Miss I. (Repression^G Rating 2, bottom rank for both R^L and R^{Lt}; Isolation^G Rating 5) in Figure 1 illustrate the points already made on the basis of correlational analyses. Each circle in Figure 1 represents a locus of fixation as it appeared on the motion picture record. The numbers within each locus represent the number of frames for each fixation, i.e., duration per fixation. Each frame of film is equal to one-sixteenth of a second. Sequence is readily read off, since the film frames are in sequence. Figure 1 shows scores on our measures for picture Number 7. It is easy to see from the picture as well as from these scores that Miss I. has more fixations and a wider dispersion of fixations than Mrs. R. The following is the recall by each subject given 8 to 10 minutes after the picture was viewed:

Recall for Mrs. R.

Mrs. R. recalled spontaneously only six of the pictures. She then was prompted on the remainder and was able to recall them eventually. The following dialogue between the experimenter and the subject occurred in recalling picture Number 7:

Experimenter (E): (giving the first cue after the subject had not been able to recall this picture) A man reading near a nude woman. Do you remember that one?

Subject (S): No. I don't remember that one at all.

E: Take a moment. A man reading near a nude woman. You don't remember that one at all?

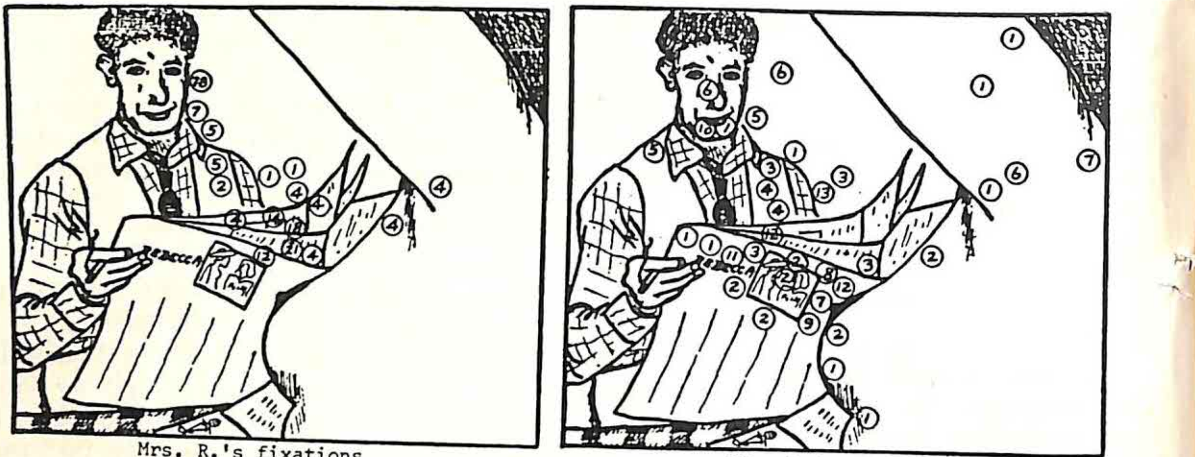
S: No, I don't remember that one at all.

E: A man seated reading a paper.

S: I remember the window. (The picture contains no window.)

E: And you could see her profile. You could see her breast.

S: Yes. Now I remember. Just the top. The girl's face turned around, no waistline, the man reading a paper, and that's all I saw.



Mrs. R.'s fixations
(High Repression)

Miss I.'s fixations
(High Isolation)

	Mrs. R.	Miss I.
"Ground" score (sec.)	0.06	0.62
Dispersion from center	0.07	0.31
Scatter	7.00	19.00
Scatter per fixation	0.41	0.59
Track length (mean in mm.)	29.50	38.10
Duration per fixation (mean in sec.)	0.68	0.28
% Sexual items recalled	picture forgotten	100.00
GSR (Rank within each S's 10 pictures)	2.00	3.00

FIG. 1. Illustration and sample scores of two women's looking behavior on Picture Number 7. (This is a sketch of the photograph used as the stimulus. Circled numbers denote position and number of frames of film for each fixation.)

E: That's fine.

(Even after several clues she had difficulty recalling the picture, and when she did she said nothing about the sexual aspects. Mrs. R.'s descriptions were typically short and vague. She had lived in this country several years but expressed herself haltingly and was insecure about her speech, although she spoke English quite understandably. She showed some support-seeking from the experimenter, while being timorous and fearful of looking at him.)

Recall for Miss I.

> *S:* This is a man reading a newspaper. There's a woman by the table with her naked bosom. I thought it was very humorous because he wasn't looking at her at all. And this would give an idea that he didn't see, hardly, a nude, especially of the opposite sex. I thought it was very humorous. He seemed unemotional and didn't seem to notice the whole thing at all. Do you want me to speak pictorially and say I saw the chair was there, or something?

(Miss I. is an attractive, vibrant young woman who has been trained in the arts. Her recalls of the pictures tended to be detailed, with no shying away from the sexual aspects—in fact, an enjoyment of them, in contrast to Mrs. R.'s neglect of them. While Mrs. R. was quiet and compliant during the experiment, Miss I. was more restive and at times even oppositional.)

CONCLUSIONS

Our main focus was on looking behavior as a function of defense. Here we discovered something new and clear: the defense of isolation is associated with venturing to look around more and the defense of repression with looking around less. Four looking measures show the trend most strikingly—the ground score, scatter, scatter per fixation, and track length. The isolators tend to look at the ground as well as the figure; they scatter their fixations and they make big jumps from one fixation to the next. The repressors, on the other hand, tend to do the opposites of these (although these correlations miss significance for Repression^G).[9] Repression^L and Repression^{Lt} give higher correlations than Repression^G here and in most comparisons throughout this paper.

We also looked over the correlates of defense with recall of pictures. The main finding here was that the defense of isolation was better than repression for recall of sexual content.[10]

---

[9] Most correlations for looking measures with the Schimek ratings of defense were higher for Isolation^G than for Repression^G. Three possibilities might explain this: (*a*) we had a smaller range of ratings in our sample for the defense of repression than for isolation; (*b*) the judge was able to estimate isolation better than repression; (*c*) the correlations represent the true state of affairs for the Gardner method: isolation ratings give higher correlations with looking measures.

No strong trend emerged from the correlations of defense ratings with GSR measures. The findings suggested that the higher the repression, the greater the GSR peak while looking at a picture.

Some looking measures, especially scatter per fixation, scatter, and track length, correlate significantly with standard Rorschach scores.

The main limit in the experiment—the moderate size of the sample—is partly compensated by the consistency and clarity of these results with two independent measures of defense, with several measures of looking behavior, and by the agreement with findings of other experimenters. In future work we will increase our variety of subjects, and explore other definitions of defense. Our present definition of repression versus isolation probably overlaps considerably with the clinical categories, hysterical versus obsessive-compulsive. It may also overlap with the remarkable dichotomy, field-dependent versus field-independent (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). The main result regarding looking behavior and defense seems consistent with some of the results reported for several cognitive styles—people whose main defense is repression tend to level, that is, to diminish stimulus differentiation (Holzman & Gardner, 1959). Perhaps not venturing to look around is an aspect of leveling behavior. Isolators were reportedly more prone to use the cognitive style of scanning (Gardner et al., 1959; Gardner & Long, 1962). Our discovery that heavy reliance upon isolation is associated with venturing to look around while taking in information is probably in agreement with these findings from the cognitive style work.

The findings on looking and recalling associated with different defenses also may be consistent with our clinical understanding of the defenses. For a long time clinicians have believed that repressors are less able than isolators to present the information from a sexual stimulus via direct recall. Some of the results on figure-ground, the scatter measures, and track-length sound remarkably like the embodiment in looking behavior of the observation (Schafer, 1948) that awareness is narrowed by repression and broadened by isolation. Looking around more is probably typically associated with greater awareness of a stimulus.[11]

Information about differences between peripheral and foveal vision suggests another difference between the looking behavior of repressors and isolators. Thomas (1963) mentions that the area covered by foveal vision subtends only two to three degrees of solid angle. In our experiment, in which the subject's eye is 28 inches from the screen, the area of foveal vision must be a total of a circle 1 inch in diameter. Since repressors, compared to isolators, look around so little, they must take in much more of the picture through peripheral vision. Peripheral vision is known to be poor for seeing details. Even when a repressor and an isolator present the same content in direct recall, the repressor may have taken it in through peripheral vision, and hence have a relatively less distinct impression.

---

[10] Byrne & Sheffield (1965) selected 44 repressors and 44 sensitizers by scores on a Repression-Sensitization (*R-S*) scale. Half of the subjects read vivid sexual passages and half read neutral passages. Sex arousal was significantly greater in the experimental group. As predicted, sensitizers were more anxious than repressors in the sex arousal condition, but the two groups did not differ in the neutral situation. Among sensitizers, sex arousal was associated with feelings of anxiety, entertainment, and lack of boredom; among repressors it was associated with disgust and anger. The authors interpret their findings to mean that sexual stimuli are equally threatening to these two defense types, but that sensitizers meet the threat by verbalizing their feelings of anxiety. Sensitizers show awareness of their state by indicating the simultaneous presence of feelings of sexual excitement, anxiety, and pleasure. These differences between repressors and sensitizers in response to sexual material by subjects of Byrne and Sheffield *may* correspond to the differences we found for repressors and isolators.

[11] In an unpublished experiment we (Luborsky, L., Rice, R., Phoenix, D., Fisher, C.) answered a related question affirmatively: Does the first clumping of fixations in an area on a picture occur at the moment of awareness of something in that area, or before or after that moment? We arranged for a projector to project a word on a screen. Since the room was bright and the image faint, the word could not be seen. Stepwise increases of illumination were made till the person could report "something," while photographs of eye movements were taken at each step. We were able to demonstrate the great simultaneity of the first clumping of fixations in the area of the word with the first discrimination of something in the visual field.

It would be good to be able to test the limits of these repression- and isolation-related behaviors. To what extent are these looking behaviors a product of our instruction "You are free to look around where you wish?" Another set might produce quite different looking behavior, even in extreme repressors and isolators. Altered task instructions were shown in a few illustrations by Yarbus (1961) to alter looking behavior significantly. The fact that we got some high correlations between behaviors on our TATlike pictures and Rorschachs would suggest stable styles of looking; however, the instructions in our task and the Rorschach task are both on the side of encouragement to look freely and both probably involve an attempt to extract information to be recalled, even though this is not explicitly asked for. A necessary future experiment would therefore be to try extreme repressors and isolators with different instructions. What would happen if we instructed subjects to look around while looking at the pictures? Would the forgetting of sexual content by repressors diminish with these instructions? Similarly, would the isolators forget more if they were instructed to fixate only on one area of the picture? Such an experiment ought to help distinguish peripheral versus central factors in the recall of the pictures. We might find that when instructions prevent a repressor from using eye fixations as part of an avoidant defense, he can easily substitute other methods of defense.

REFERENCES

BUSWELL, G. T. *How people look at pictures: A study of the psychology of perception in art.* Chicago: Univer. Chicago Press, 1935.

BYRNE, D. The repression-sensitization scale: Rationale, reliability, and validity. *Journal of Personality,* 1961, **29**, 334–349.

BYRNE, D., & SHEFFIELD, J. Response to sexually arousing stimuli as a function of repressing and sensitizing defenses. *Journal of Abnormal Psychology,* 1965, **70**, 114–118.

GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, HARRIET, & SPENCE, D. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues,* 1959, **1**, No. 4.

GARDNER, R. W., & LONG, R. I. Control, defense and centration effect: A study of scanning behavior. *British Journal of Psychology,* 1962, **53**, 129–140.

HOLZMAN, P. S. Repression and cognitive style. *Bulletin of the Menninger Clinic,* 1962, **26**, 273–282.

HOLZMAN, P. S., & GARDNER, R. W. Leveling and repression. *Journal of Abnormal and Social Psychology,* 1959, **59**, 151–155.

LEVINE, M., & SPIVACK, G. The Rorschach index of ideational repression: Application to quantitative sequence analysis. *Journal of Projective Techniques,* 1963, **27**, 73–78.

LEVINE, M., & SPIVACK, G. *The Rorschach index of repressive style.* Springfield, Ill.: Charles C Thomas, 1964.

LUBORSKY, L. Momentary forgetting during psychotherapy and psychoanalysis: A theory and research method. In R. R. Holt (Ed.), *Contributions to a psychoanalytic theory of behavior: Memorial essays in honor of David Rapaport.* New York: International Universities Press; in press.

LUBORSKY, L., BLINDER, B., & MACKWORTH, N. Eye fixation and recall of pictures as a function of GSR responsivity. *Perceptual & Motor Skills,* 1963, **16**(5-V16), 469–483.

LUBORSKY, L., BLINDER, B., & MACKWORTH, N. Eye fixation and the contents of recall and images as a function of heart rate. *Perceptual & Motor Skills,* 1964, **18**, 421–436.

LUBORSKY, L., & SHEVRIN, H. Forgetting of tachistoscopic exposures as a function of repression. *Perceptual & Motor Skills,* 1962, **14**, 189–190.

MACKWORTH, N. Some suggested uses for the Optiscan—a head-mounted eye camera. Paper read at the American Society of Mechanical Engineers, November–December 1960. Pp. 1–7.

MACKWORTH, N., & KAPLAN, I. Points of view and lines of sight. *Ikon,* 1963, **13**, 45–60.

MACKWORTH, N., & THOMAS, E. L. Head-mounted eye-marker camera. *Journal of the Optical Society of America,* 1962, **52**, 713–716.

SCHAFER, R. *The clinical application of psychological tests.* New York: International Universities Press, 1948.

SHANNON, D. T. Clinical patterns of defense as revealed in visual recognition thresholds. *Journal of Abnormal and Social Psychology,* 1962, **64**, 370–377.

SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *Journal of Abnormal and Social Psychology,* 1958, **56**, 285–294.

THOMAS, E. L. Eye movements and fixations during initial viewing of Rorschach cards. *Journal of Projective Techniques and Personality Assessment,* 1963, **27**, 345–353.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

YARBUS, A. L. Eye movements during the examination of complicated objects. *Biophysics* (USSR), 1961, **6**, 52–56.

(Received June 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 281–284

"DEFENSE" AGAINST TRAUMATIC CONCEPTS[1]

ELI SALTZ AND ADOLPH DI LORETO

Center for the Study of Cognitive Processes, Wayne State University

The present study shows that traumatic responses attached to 1 instance of a concept will impair the identification of the concept even when the traumatized concept is not part of the concept identification task. Such traumatic priming proved detrimental compared with either neutral priming or no priming. Neutral priming per se proved to facilitate concept identification.

Saltz (1961, 1963) has proposed that a crucial element of cognitive systems is the ability to resist interference from competing systems (i.e., systems can be conceptualized as having "boundary strength"). In a recent series of studies (Saltz, 1961; Saltz & Asdourian, 1963; Saltz & Riach, 1961) data have been presented which indicate that stress induced external to a system (e.g., shock induced while the subject is responding to a visual discriminandum) breaks down the boundaries of such a system. The present study is concerned with examining the effect of stress on boundaries when the stress is *internal,* within the system.

Staats, Staats, and Heard (1959) have shown that if an instance of a concept is associated with affectively loaded words, the affect generalizes to other instances of the concept. Thus, if an instance of a previously neutral concept is associated with traumatic material, the traumatic affect becomes a part of the concept system: the entire concept takes on some degree of the affect. The previously cited work on the effect of stress on boundary strength leads to the prediction that stress within a concept system should weaken the boundaries of the system. Consequently, it should be difficult for the subject to identify the relevant concept.

To insure that the phenomenon was to the generalized concept, rather than to the concept instance directly associated with traumatic stimuli, the traumatized instance was not included in the critical test of concept identification.

METHOD

Subjects

The subjects were 90 undergraduates at Wayne State University, 50 assigned to Experiment 1, 40 to Experiment 2. In Experiment 1, 22 subjects were male, 28 female. In Experiment 2, 22 subjects were male, 18 female. Two lists were employed in each experiment; half the subjects in each experiment were assigned to each list.

Materials

The first 10 items of a test developed by Rosen (1962), the Verbal Concept Attainment Test, were used as the concept identification task. Each item consisted of six rows of words, four words per row. One word in each row was an instance of the correct concept for the item. Table 1 illustrates an item in which the correct concept is "animal." (In the actual test, of course, the correct instances were not in capital letters.) Each item was presented on a separate 3 × 5 index card. Score on an item was the total time required for the subject to indicate which six words, one in each row, were instances of a single concept. The same 10 items were used in both Experiment 1 and Experiment 2.

In Experiment 1, a paired-associates learning task was used to attach traumatic or neutral responses to a seventh instance for each of the 10 concepts to be used in the concept identification task. Two lists were prepared, both containing the following stimulus terms: ORCHID, EAR, TUBA, HORSE, STUDY, BALLET, CAR, HUGE, PLUMBER, LOVE. Five of these stimuli were associated to the following traumatic responses: WHORE, SYPHILIS, RAPE, BITCH, VOMIT. The remaining five stimuli were associated with the following neutral responses: GLASS, SIDE, ELEVEN, WITH,

TABLE 1

SAMPLE CONCEPT IDENTIFICATION ITEM

READ	PARTY	*TIGER*	PEARLS
BOOK	*MONKEY*	FEMALE	THRILL
LEAP	PAPER	TEETH	*DONKEY*
MOUTH	*WHALE*	SHOE	PLAY
SNAKE	MUSIC	COUCH	SILLY
FLOOR	PICTURE	NECK	*HAWK*

---

[1] The present study was supported in part by National Science Foundation Grant GB 2538.

TABLE 2

Mean Time (in Seconds) to Identify Concepts when Instances of the Concepts Have Been Associated with Traumatic or Neutral Responses

	Traumatic		Neutral	
	M	*SD*	*M*	*SD*
Material 1	30.8	14.4	25.1	9.2
Material 2	24.4	11.4	21.6	7.3
Grand *M*	27.6		23.3	

Note. The stimuli associated with traumatic responses in List 1 were associated with the neutral responses in List 2 and vice versa.

The material of Experiment 2 differed from the material of Experiment 1 only in the paired-associates lists learned prior to the concept attainment task. The lists differed in two respects. In Experiment 1, half the responses in a list were traumatic, half neutral. In Experiment 2, all 10 responses in each list were neutral. The second difference was in the relationship between the stimuli of the paired-associates lists and the concepts in the concept attainment task. In Experiment 1, each of the 10 stimuli in a paired-associates list was an instance of a concept in the concept attainment task. In Experiment 2, only 5 stimuli of a list were instances of the concepts. The other 5 stimuli in a list were buffer items. Thus, a subject in Experiment 2 was primed for only 5 of the concept problems he was to subsequently solve, and was not primed for the other 5 concept problems.

Experiment 2 employed two paired-associates lists. Both lists contained the same 10 neutral response terms: CLOCK, SILLY, NOTE, STRING, SIDE, WORD, GLASS, WITH, ELEVEN, BOOK. (Five of these are the neutral response terms from Experiment 1.) List 1 of Experiment 2 used as stimuli 5 of the stimuli employed in Experiment 1 (i.e., ORCHID, EAR, TUBA, HORSE, STUDY) plus 5 new stimuli which were irrelevant to the subsequent concept identification task (i.e., RADIO, WHEEL, DREAM, TABLE, RICH). List 2 employed the remaining 5 stimuli from Experiment 1 which were not in List 1, plus the same 5 irrelevant stimuli from List 1.

Procedure

Each subject was tested individually. First, the subject learned his appropriate paired-associates list for nine trials on a Lafayette memory drum. Each trial consisted of a learning phase in which the pairs were presented at a 2-second rate followed (6 seconds later) by a test phase in which the stimuli were presented alone and the subject had 2 seconds to give the correct response before the next stimulus appeared. Three different random orders of pairs were used for the learning phases and a different three random orders for the test phases.

Learning was followed by a 2-minutes rest interval before the onset of the concept identification task. The subjects were presented with each concept item for a maximum of 3 minutes. (None of the subjects required the full 3 minutes for any problem.) Time for solution was individually determined for each subject. The 10 items were presented in a different random order to each subject so as to randomize practice effects. Instructions were to find the word in each row which, together with one word from each of the other rows, would form a familiar verbal concept. One sample item was presented as an illustration before the presentation of the 10 critical items.

Results

Experiment 1

The results of the present experiments clearly show that learning traumatic responses to instances of a concept produces difficulty in identifying the concept. Table 2 summarizes the data from Experiment 1. Since each subject was tested on concepts to which he had associated traumatic responses and on concepts to which he had associated neutral responses, the data were suitable to a Latin-square analysis of variance with repeated measures. The difference in time to solve traumatized concepts as opposed to neutral concepts was significant beyond the .01 level ($F = 7.46$, $df = 1/48$). While the two sets of materials differed significantly in difficulty ($F = 9.78$, $df = 1/48$), the effect of trauma was approximately equal for the two sets of materials (i.e., the interaction with material did not approach significance, $F = 0.32$, $df = 1/48$).

Experiment 2

Having the subjects learn responses to instances of a concept may produce a priming effect, facilitating identification of the concept, even when the instances involved in learning are not among those involved in the concept identification task. Consequently, it is possible that the traumatic priming in Experiment 1, while producing poorer concept identification than the neutral priming, might actually be facilitory compared with a condition involving no priming. Experiment 2 was designed to cast light on this problem.

In Experiment 2, subjects learned neutral responses either to the same concept instances involved in Experiment 1 or to completely irrelevant stimuli which would not prime the concepts in the concept identification phase. Across the two sets of materials in Experiment 2, the mean time to solve the neutral primed concepts was 18.7 seconds, the mean time to solve nonprimed concepts was 22.2 seconds. The effect of priming was significant well beyond the .05 level ($F = 6.65$, $df = 1/38$) and the interaction between priming and material failed to reach significance ($F < 1.00$).

Comparison of the mean concept solution time following traumatic priming in Experiment 1 with that for the nonprimed condition of Experiment 2 (27.6 seconds versus 22.2 seconds) shows that the traumatic priming produces an absolute decrement in performance. To evaluate this difference, a 2 × 2 factorial design was executed (Traumatic Priming versus Nonpriming × Material 1 versus Material 2). Compared with the nonprimed condition of Experiment 2, the decrement in performance for the traumatic priming of Experiment 1 was significant well beyond the .05 level ($F = 5.887$, $df = 1/86$). This effect did not interact significantly with material ($F = 1.558$, $df = 1/86$).

Comparison of Experiment 1 with Experiment 2 shows that the mean time for solution of the concepts across material is slower for the neutral priming of Experiment 1 than for the neutral priming condition of Experiment 2 (23.3 seconds versus 18.7 seconds). Since the concepts to be solved are identical for the two conditions, as are the instances and their responses during initial learning, the difference must be due to the fact that the subjects in Experiment 1 learned traumatic responses to half of the stimuli. The decrement in concept identification performance for the neutral stimuli of Experiment 1 appears to indicate a *generalized* effect of the traumatic responses. A 2 × 2 analysis of variance (Neutral Priming Experiment 1 versus Experiment 2 × Material 1 versus Material 2) indicates that the decrement in performance for the neutral condition in Experiment 1 is significant beyond the .01 level ($F = 7.638$, $df = 1/86$). This effect did not interact significantly with material ($F = 1.086$, $df = 1/86$).

Turning to the paired-associates learning data, we find that traumatic and neutral responses in Experiment 1 were learned at approximately the same rate, as can be seen in Table 3. Analysis of variance (Latin square for repeated measures) indicated no significant differences between traumatic and neutral items ($F = 2.994$, $df = 1/48$), between the two sets of materials employed ($F = 2.46$, $df = 1/48$) or for the interaction term ($F = 1.79$, $df = 1/48$). It is interesting to note that Kelley (1964), in a study which carefully controlled Thorndike-Lorge (1944) frequency for neutral and traumatic materials, as well as associative frequency with the relevant stimulus terms as determined by the Russell and Jenkins (1954) word association norms, also failed to find any difference in paired-associates learning of traumatic versus neutral responses. Consequently, the difference in Experiment 1 between Traumatic and Neutral conditions in concept identification cannot be accounted for by fewer reinforcements, during learning, of the traumatized stimuli.

Finally, an examination of sex differences indicates that women are somewhat faster than men at solving the concept problems in all conditions. Also, women seem somewhat less influenced than men by either traumatic or neutral priming. On the other hand, a careful examination of the data indicates that all the effects reported as significant (i.e., Traumatic versus Neutral; Primed versus Unprimed; Traumatic versus Unprimed; Neutral Primed Experiment 1 versus Experiment 2) are characteristic of both men and women.

DISCUSSION

The present study indicates that stress, produced by a traumatic response being asso-

TABLE 3

MEAN NUMBER OF CORRECT RESPONSES FOR INITIAL LEARNING OF TRAUMATIC AND NEUTRAL MATERIAL OVER THE NINE LEARNING TRIALS

	Traumatic		Neutral	
	M	*SD*	*M*	*SD*
Material 1	37.8	4.71	36.9	5.59
Material 2	35.3	3.98	37.6	4.23
Grand *M*	36.6		37.2	

ciated to an instance of a concept, can generalize to the concept per se. In addition, such stress can reduce the efficiency of identifying the concept. The results are consistent with the theoretical position proposed by Saltz (1961) which states that stress reduces the ability of a cognitive system to withstand interference from competing systems. One of the important characteristics of the present study is that stress was *internal,* within the system.

Certain gross similarities can be seen between the present concept identification task and the "perceptual defense" type of task (e.g., Spence, 1957). In the perceptual defense studies, subject's threshold is determined for identifying traumatic versus neutral stimuli. In the concept identification situation, all the stimuli are suprathreshold, but the relevant concept must be identified. The perceptual defense situation can be conceptualized as one in which the total hierarchy of subject's response systems compete with the traumatized system to be perceived. Thus, the writers' interpretation of the present results could also be applicable to perceptual defense.

It is important to emphasize that the writers' interpretation of the traumatized concept problem does not involve avoidance behavior. Instead, the crucial factor is an alteration of one of the basic elements of the concept system, namely the "boundary strength" of the concept. An avoidance hypothesis, such as those often encountered in the perceptual defense area, would assume that subjects detected the traumatized stimuli at some subliminal level, but then avoided them. The writers' formulation makes it unnecessary to assume such detection processes. Consequently, the writers' position can account for related findings, such as those of Saltz and Riach (1961) in which stress produced decrements in discrimination performance despite the fact that the stress was general and not attached to the correct response.

REFERENCES

KELLEY, R. J. The learning and recall of emotional and neutral words in negative transfer situation as a function of differentiation. Unpublished doctoral dissertation, Wayne State University, 1964.

ROSEN, H. *Verbal concept attainment test.* Detroit: Center for the Study of Cognitive Processes, Wayne State University, 1962.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosenoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr 66216, Office of Naval Research.

SALTZ, E. The effect of induced stress on free associations. *Journal of Abnormal and Social Psychology,* 1961, **62**, 161–164.

SALTZ, E. Differentiation as a cognitive process. In Irving Sigel (chm.), Structure and Process in Cognitive Theory. Symposium presented at American Psychological Association, Philadelphia, 1963.

SALTZ, E., & ASDOURIAN, D. Incubation of anxiety as a function of cognitive differentiation. *Journal of Experimental Psychology,* 1963, **66**, 17–22.

SALTZ, E., & RIACH, W. J. The effect of stress on stimulus differentiation. *Journal of Experimental Psychology,* 1961, **62**, 588–593.

SPENCE, D. P. A new look at vigilance and defense. *Journal of Abnormal and Social Psychology,* 1957, **54**, 103–108.

STAATS, A. W., STAATS, C. K., & HEARD, W. G. Language conditioning of meaning using a semantic generalization paradigm. *Journal of Experimental Psychology,* 1959, **57**, 187–192.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received July 6, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 285–289

SCHOOL PHOBIA:

RAPID TREATMENT OF FIFTY CASES[1]

WALLACE A. KENNEDY

Human Development Clinic, Florida State University

The Human Development Clinic of Florida State University has determined that a differential diagnosis between 2 types of school phobia can be made on the basis of any 7 of 10 differential symptoms. The rapid treatment program for Type 1, the neurotic crisis type of school phobia, is presented in 6 steps which extend over a 3-d. period. All 50 cases over the past 8 yrs responded to the treatment with a complete remission of school phobia symptoms. Follow-up study indicates no evidence of any outbreaks of substitute symptoms.

School phobia, a dramatic and puzzling emotional crisis, has attracted considerable attention for a number of years. Phobias in general are the subjects of widely differing theories of dynamics and treatment. The controversy regarding the treatment of children's phobias dates from the earliest case studies presented by Freud (1909), continues through the laboratory demonstrations of Watson and Jones (Jones, 1924) to the more recent experimental treatment of Wolpe (1954). There have been five broad reviews since the earliest paper presented by Johnson et al. in 1941: Klein (1945), Waldfogel, Cooledge, and Hahn (1957), Kahn (1958), Glasser (1959), and Sperling (1961). These reviews in the main support the contention that the major weight of evidence thus far leans toward the psychoanalytic interpretation of phobias, while the work of Wolpe is more consistent with the approach presented herein.

The psychoanalytic theory stresses the role of the mother in the development of school phobia. A close symbiotic relationship, which displays itself in an overdependency, is present between the mother and child. Stemming from an unsatisfactory relationship with her own mother, the mother finds it difficult to cope with her own emotional needs. The father often is in a competing role with the mother, and seems to try to outdo her in little tasks around the home: in trying to strengthen his own image, he depreciates that of the mother. He too overidentifies with the child. Thus, the emotional climate of his family prevents the child from ever finding out whether or not he, of his own volition, can solve problems. Possessive, domineering parents tend to make the child's growth toward independence difficult. His guilt regarding his own impulses is transformed into depression: the anxiety can reach extreme proportions.

On the other hand, Wolpe sees the phobia as a learned reaction, which he treats through direct symptom attack with what he calls reciprocal inhibition, or desensitization.

Interest in the school phobia problem, which occurs at the rate of 17 cases per thousand school-age children per year, has been greatly intensified in the past few years. An extremely significant advance was made by Cooledge and the Judge Baker group in 1957, when they presented evidence that there were not one, but two types of school phobia, which, although sharing a common group of symptoms, differed widely in others. These are referred to as Type 1 School Phobia, or the neurotic crisis, and Type 2 School Phobia, or the way-of-life phobia. The common symptoms are: (*a*) Morbid fears associated with school attendance; a vague dread of disaster; (*b*) Frequent somatic complaints: headaches, nausea, drowsiness; (*c*) Symbiotic relationship with mother, fear of separation; anxiety about many things: darkness, crowds, noises; (*d*) Conflict between parents and the school administration.

At the Human Development Clinic of Florida State University, 10 differential symp-

---

[1] The research reported herein was supported through the Human Development Clinic of Florida State University, Tallahassee, Florida.

toms between Type 1 and Type 2 School Phobia have been determined. A differential diagnosis can be made logically and empirically on the basis of any 7 of the 10.

Ten Differential School Phobia Symptoms

Type 1	Type 2
1. The present illness is the first episode.	1. Second, third, or fourth episode.
2. Monday onset, following an illness the previous Thursday or Friday.	2. Monday onset following minor illness not a prevalent antecedent.
3. An acute onset.	3. Incipient onset.
4. Lower grades most prevalent.	4. Upper grades most prevalent.
5. Expressed concern about death.	5. Death theme not present.
6. Mother's physical health in question: actually ill or child thinks so.	6. Health of mother not an issue.
7. Good communication between parents.	7. Poor communication between parents.
8. Mother and father well adjusted in most areas.	8. Mother shows neurotic behavior; father, a character disorder.
9. Father competitive with mother in household management.	9. Father shows little interest in household or children.
10. Parents achieve understanding of dynamics easily.	10. Parents very difficult to work with.

PROBLEM

In the Fall of 1957 the Clinic embarked upon an experimental procedure for the treatment of Type 1 School Phobia—a procedure similar to that of Rodriguez, Rodriguez, and Eisenberg (1959) with one major exception: whereas Rodriguez made no distinction between types of school phobia and treated in the same manner all cases which came to the clinic, the 50 cases reported herein were selected on the basis of the criteria mentioned above. The Florida State University Human Development Clinic, as a teaching and research clinic, does not generally see deeply disturbed children, but refers them to other agencies.

In the 8-year period covered by the report, there have been 6 cases which would meet the criteria of Type 2 School Phobia. These 6 cases were treated by supportive therapy for the children and parents. None of the 6 Type 2 cases had more than three of the 10 Type 1 criteria, and the results were completely dissimilar to those reported for the 50 Type 1 cases. All of the Type 2 cases were chronic in nature. All had family histories of one or more parents seriously disturbed. Two of the cases were diagnosed as having schizophrenia; 2 were diagnosed as having character disorders with the school phobia being a minor aspect of the case. One of the 6 was hospitalized; 1 was sent to a training school. Of the 4 remaining, 2 were able to go to college, although their records were poor and their symptoms continued. These 6 cases were in treatment for an average of 10 months. In no circumstances was a school phobia case changed from Type 1 to Type 2, or vice versa.

This experimental procedure with Type 1 School Phobia was begun with considerable caution, with only 1 case in 1957 and 2 the following Spring. The treatment involved the application of broad learning theory concepts by blocking the escape of the child and preventing secondary gains from occurring. In addition, the child was reinforced for

TABLE 1

YEAR OF TREATMENT AND SEX OF 50 TYPE ONE SCHOOL PHOBIA CASES

Year	Male	Female	Total
1957	1	0	1
1958	1	1	2
1959	4	2	6
1960	4	8	12
1961	6	3	9
1962	5	4	9
1963	4	5	9
1964	0	2	2
Total	25	25	50

TABLE 2

SYMPTOM CHECKLIST AND SEX OF 50 TYPE ONE SCHOOL PHOBIA CASES

Symptom	Male	Female	Total
1. First attack	25	25	50
2. Monday onset—Thursday illness	24	25	49
3. Acute onset	25	23	48
4. Lower grades	22	18	40
5. Death theme	22	22	44
6. Mother's health an issue	23	21	44
7. Good parental marital harmony	24	23	47
8. Good parental mental health	23	24	47
9. Father helper in the house	21	21	42
10. Parents achieve insight quickly	24	25	49

TABLE 3

Age and Sex of 50 Type One School Phobia Cases

Age	Male	Female	Total
4	0	1	1
5	3	1	4
6	2	3	5
7	3	2	5
8	3	1	4
9	3	5	8
10	4	3	7
11	1	2	3
12	3	0	3
13	2	4	6
14	1	2	3
15	0	0	0
16	0	1	1
Total	25	25	50

going to school without complaint. This rapid treatment procedure has now been followed with 50 cases.

Subject Population

Subjects for the 50 cases over an 8-year period were school-age children, all suffering from the first evidence of a phobic attack, from the geographical area served by the Human Development Clinic of Florida State University. The subject distribution by year and sex is illustrated in Table 1, by symptom and sex in Table 2, by age and sex in Table 3, and by grade and sex in Table 4.

The fathers' mean age for the male subjects was 36; the mothers', 35. For the female subjects the fathers' mean age was 38; the mothers', 36. The boys' mean age was 9; that of the girls', 10. There was no definite pattern in birth order of the subjects, or in number of siblings.

Method and Results

During the course of the past 8 years, 50 cases of Type 1 School Phobia have been treated. Five of these cases might be considered semicontrols because they were untreated Type 1 cases of some duration, or they were Type 1 cases unsuccessfully treated elsewhere before they were seen at the clinic. One of these semicontrol cases had been out of school for 1 year, and the other 4 had been out for over 3 months.

All 50 of the cases responded to the treatment program with a complete remission of the school phobia symptoms, and follow-up study indicates no evidence of any outbreaks of substitute symptoms or recurrence of the phobia.

In the follow-up schedule the parents were phoned in about 2 weeks, and again in 6 weeks, to see if the progress had continued. They were then phoned on a yearly basis, except in 1961, when follow-up interviews were conducted reaching 19 of the 21 cases completed at that time. During the course of the 8 years, 6 families were lost because of moving with no forwarding address. Of these lost cases, none had been followed less than 2 years, 2 were followed 3 years, and 1 for 4 years.

Rapid Treatment Procedure

The rapid treatment program for Type 1 school phobia involves six essential components: good professional public relations, avoidance of emphasis on somatic complaints, forced school attendance, structured interview with parents, brief interview with child, and follow-up.

Good Professional Public Relations

It is necessary to establish good communication with schools, physicians, and parent groups, such that the cases are likely referred on the second or third day of the phobic attack. This groundwork involves the typical mental health consultation and case-by-case follow-up with the referring source.

Avoidance of Emphasis on Somatic Complaints

If phobic qualities predominate, that is, if the child conforms to seven of the differential symptoms of Type 1 School Phobia, emphasis on somatic complaints should be avoided. For instance, the child's somatic complaints should be handled matter-of-factly, with an appointment to see the pediatrician after school hours. Abdominal pains will probably require the pediatrician to make a prompt physical examination, but this can probably be done on the way to school.

TABLE 4

Grade and Sex of 50 Type One School Phobia Cases

Grade	Male	Female	Total
Nursery School	0	2	2
Kindergarten	4	0	4
First	4	4	8
Second	0	1	1
Third	6	4	10
Fourth	3	4	7
Fifth	2	2	4
Sixth	3	1	4
Seventh	2	2	4
Eighth	0	2	2
Ninth	1	2	3
Tenth	0	1	1
Total	25	25	50

Forced School Attendance

It is essential to be able to require the child to go to school and to be willing to use any force necessary. In all of the present cases, simply convincing the parents of this necessity and having them come to a firm decision, has generally been enough. The ability to be decisive when necessary has been essential.

Have the father take the child to school. These fathers are not unkind, and they can show authority when necessary.

Have the principal or attendance officer take an active part in keeping the child in the room.

Allow the mother to stand in the hall, if she must, or to visit the school during the morning, but not to stay.

Structured Interview with the Parents

Stressing the following points, conduct with the parents a structured interview designed to give them sufficient confidence to carry out the therapeutic program even in the face of considerable resistance from the child.

Lead the interview. The confidence of the parents is greatly increased by the interviewer's verifying the history rather than taking it. Correctly anticipating 7 out of 10 variables within a family structure is well calculated to induce full cooperation.

Be optimistic. Stressing the transient nature, the dependable sequence of a difficult Monday, a somewhat better Tuesday, and a symptom-free Wednesday, tends to lighten the depression of the parents regarding their child's unwillingness to go to school.

Emphasize success. Type 1 cases always recover. Ninety percent of the Type 1 phobics stay at school most of the first day. Along with optimism comes a slight mobilization of hostility which helps the parents to follow the plan.

Present the formula. Simply but directly, with repetition for emphasis, outline a plan for the parents to follow, assuming that it is the end of the school week by the time of the referral and that the interview with the parents is conducted on Thursday or Friday.

Parent Formula

Do not discuss in any way, school attendance over the weekend. There is nothing a phobic child does better than talk about going to school. Don't discuss going to school. Don't discuss phobic symptoms. Simply tell the child Sunday evening, "Well, son, tomorrow you go back to school."

On Monday morning get the child up, dressed, and ready for school. Give the child a light breakfast to reduce the nausea problem. Have the father take the child matter-of-factly off to school. Don't ask him how he feels, or why he is afraid to go to school, or why he doesn't like school. Simply take him to school, turn him over to the school authorities, and go home.

If the child therapist has not seen the child the previous week, he may see him after school on the first day.

On Monday evening, compliment the child on going to school and staying there, no matter how resistant he has been, no matter how many times he has vomited, cried, or started to leave. If he has been at school for 30 minutes on Monday, progress is being made. Tell the child Monday evening that Tuesday will be much better, and make no further mention of the symptom.

Tuesday can be expected to be a repetition of Monday, but with everything toned down considerably. On Tuesday evening, encourage and compliment the child strongly for doing so much better.

Wednesday should be virtually symptom free. Wednesday evening, with considerable fanfare, give a party for the child in honor of his having overcome his problem.

Brief Interview with the Child

The child himself should be seen only briefly by the child therapist and only after school hours. The content of the interview should be stories which stress the advantage of going on in the face of fear: how student pilots need to get back into the air quickly after an accident, and how important it is to get right back on the horse after a fall. In addition the therapist can describe real or imaginary events in his own childhood when he was frightened for awhile but everything turned out all right: all to stress to the child the transitory nature of the phobia.

Follow-Up

Follow-up by phone, being chatty and encouraging and not oversolicitous. In the long-range follow-up, chat with the parents about further school phobia symptoms, incidence of other phobias, school attendance records, academic progress, and the occurrence of other emotional problems in the child.

Discussion

Two legitimate concerns have been expressed regarding preliminary reports at local meetings. The first is a concern about the claim of complete remission for all 50 cases—a claim inconsistent with the usual child guidance clinic success rate—and the consequent belief that the criterion for success is simply too narrow. Only self-report data and reports from school administrations are available regarding the symptom-free nature of these children once this phobic episode has passed. It is true that no diagnostic evaluation has been undertaken with any of these children during follow-up. It must be remembered, however, that the definition of symptom remission is restricted to those obvious symptoms which might conceivably lead the parents or school officials to re-refer the children to the clinic. In this regard, these 50 children in the Type 1 School Phobia group are symptom free.

Because of the nature of the Human Development Clinic and the nature of this project, careful selection has been exercised in accepting cases, as mentioned above. Due to the relationship between the schools and the clinic, and the clear definition of cases suitable for the project, there is reason to believe that the majority of Type 1 School Phobia cases in the five-county area the clinic serves have come to our attention, whereas the local county mental health clinic has received a high percentage of the Type 2 cases. The success of the Type 2 cases of school phobia accepted by the Human Development Clinic for teaching purposes has not been remarkable.

The second concern is that perhaps what is called Type 1 School Phobia is not really a severe phobic attack at all, but borders on malingering of a transient nature which would spontaneously remit in a few days anyway. In fact, because of the apparent sound mental health of the family as a group, its middle-class values which stress school, and the family's good premorbid history, including the academic record of the child, there is little reason to doubt that the majority of the cases would eventually return to school whatever treatment was undertaken. However, our five semicontrol cases and evidence seen from other clinics of Type 1 cases that have been out of school for prolonged periods suggest that this method of treatment may accelerate or facilitate the remission. Recommendation for the use of this technique is restricted, then, to those cases showing Type 1 symptoms which, in spite of their possible transient nature, present a rather serious problem to teachers, parents, and counselors.

REFERENCES

Cooledge, J. C., Hahn, Pauline B., & Peck, Alice L. School phobia: Neurotic crisis or way of life. *American Journal of Orthopsychiatry,* 1957, **27,** 296–306.

Freud, S. *Analysis of a phobia in a five-year-old boy.* Std. Ed., New York: W. W. Norton, 1909.

Glasser, K. Problems in school attendance: School phobia and related conditions. *Pediatrics,* 1959, **55,** 758. (Abstract)

Johnson, A. M., et al. School phobia: A study of five cases. *American Journal of Orthopsychiatry,* 1941, **11,** 702. (Abstract)

Jones, M. C. A laboratory study of fear: The case of Peter. *Journal of Genetic Psychology,* 1924, **31,** 308–315.

Kahn, J. H. School refusal—some clinical and cultural aspects. *Medical Officer,* 1958, **100,** 337. (Abstract)

Klein, E. The reluctance to go to school. *Psychoanalytic Study of the Child,* 1945, **1,** 263. (Abstract)

Rodriguez, A., Rodriguez, Maria, & Eisenberg, L. The outcome of school phobia: A follow-up study based on 41 cases. *American Journal of Psychiatry,* 1959, **116,** 540–544.

Sperling, M. Analytic first aid in school phobias. *Psychoanalytic Quarterly,* 1961, **30,** 504. (Abstract)

Waldfogel, S., Cooledge, J. C., & Hahn, P. B. Development, meaning and management of school phobia. *American Journal of Orthopsychiatry,* 1957, **27,** 754. (Abstract)

Wolpe, J. Reciprocal inhibition as the main basis of psychotherapeutic effects. *A.M.A. Archive of Neurology and Psychiatry,* 1954, **72,** 204–226.

(Received July 6, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 290–293

ACQUIESCENCE AND THE MMPI:

AN ITEM REVERSAL APPROACH[1]

EDWARD LICHTENSTEIN
Southern Illinois University

AND

JAMES H. BRYAN
Northwestern University

An item-reversal technique was employed to assess acquiescence on the MMPI. Reverse items were constructed for 12 MMPI scales and the 2 forms, standard and reverse, were administered in counterbalanced order to 54 psychiatric patients and 35 normal Ss. True-keyed and false-keyed subscales were scored separately to highlight acquiescence effects. The correlations between true and false subscales and their reverse forms tended, with few exceptions, to approximate the test-retest reliabilities of these scales particularly for the normal Ss. The results were interpreted to indicate that factor analytic techniques have exaggerated the role of acquiescence on the MMPI.

Acquiescence response set refers to the tendency to agree consistently with inventory items irrespective of item content. Much attention has been paid to the role of acquiescence response set in personality inventories, particularly the Minnesota Multiphasic Personality Inventory (MMPI). It has been argued (Jackson & Messick 1961, 1962) that one of two factors consistently emerging from factor analytic studies of the MMPI can be best interpreted as reflecting acquiescence (while the other factor is interpreted as reflecting social desirability).

The evidence for acquiescence effects is derived almost exclusively from studies employing factor analytic techniques. Messick and Jackson (1961a) examined factor loadings for MMPI scales from eight separate studies. They computed rank order correlations between loadings on the largest factor in each study and the proportion of items keyed true on each scale. It was found that, "Of 11 different subject samples represented in these eight studies, significant correlations were obtained for 8 of them, four of the coefficients exceeding .85 [p. 300]."

Jackson and Messick (1961, 1962) performed three large factor analyses as a means of estimating variance attributable to acquiescence on the MMPI. In each study they factor analysed correlations among MMPI scales scored separately for true- and false-keyed items, as well as five scales measuring various level of social desirability. In all three samples two very large factors appeared to be clearly identifiable as acquiescence and social desirability. The acquiescence factor accounted for 45% of the common variance in a prison sample and 26% of the common variance in neuropsychiatric and college-student samples.

If, as Jackson and Messick contend, MMPI scores are heavily influenced by acquiescence, then interpretations of MMPI scores in terms of content are seriously undermined. This study was undertaken to assess the role of acquiescence on the MMPI by a different method than factor analysis. The method chosen was the item-reversal technique widely used in studying the degree of acquiescence variance on the California F Scale (Christie, Havel, & Seidenberg, 1958). In the item-reversal method, the subjects are presented items in standard form on one occasion and in reverse form on a second occasion. For example, the item, "I have a good appetite" in the standard form becomes, "I have a poor appetite" in the reverse form. The subject must, therefore, reverse his answers to maintain logical (content) consistency. The magnitude of the correlations between scores on standard scales and scores on reverse scales reflects the degree to which the subjects respond to item content rather than manifesting acquiescence response set.

---

[1] The authors acknowledge the help of the following in collecting and processing the data: Gary Gorlick, Gwen L. Gerber, Cathleen Briix, Iris A. Lubitz, and Leslie Leventhal. The aid of the Biostatistics Division, University of California at Los Angeles Center for the Health Sciences, in performing the computations is also gratefully acknowledged. Sarah Lichtenstein contributed helpful criticism and editorial assistance.

This study was carried out while both authors were at the Neuropsychiatric Institute, University of California at Los Angeles Center for the Health Sciences.

With one limited exception, the item-reversal technique has not been applied to MMPI scales. Chapman and Campbell (1959) found a correlation of .84 between the standard and reversed forms of the Manifest Anxiety scale (*MA* scale, Taylor, 1953), which is composed of 50 MMPI items, and concluded that the *MA* scale is free of acquiescence variance. These findings suggest the fruitfulness of applying this technique to the major MMPI scales, and comparing the results with the conclusions reached through factor analysis.

METHOD

Reverse items were constructed for 12 of the 13 major MMPI scales (excluding Social Introversion). As a general rule, items were reversed in the most literal way. The standard and reversed forms of the MMPI were administered in counterbalanced order under the usual instructions.

Two groups of subjects were tested: 35 normals (25 hospital volunteer workers and 10 college students), and 54 newly admitted psychiatric patients. The volunteer-student sample was comprised of 6 men and 29 women with ages ranging from 15 to 56 ($M = 31$); the psychiatric sample consisted of 20 men and 34 women of ages ranging from 17 to 65 ($M = 32$). The test-retest interval for the volunteer-student sample was no longer than 2 days ($M = 1.1$) and for the patient sample no longer than 4 days ($M = 1.7$).

The scales on both forms were scored separately for true-keyed and false-keyed items so as to highlight the effects of acquiescence. The correlations of the true-keyed subscales with their reverse forms, and of the false-keyed subscales with their reverse forms, constitute the major results. The scoring was such that a positive correlation between regular and reverse forms indicates that items were responded to in terms of content, while a negative correlation indicates the operation of acquiescence.

Knowledge of the test-retest reliabilities of the true and false subscales is required before the standard-reverse correlations can be interpreted. Test-retest MMPIs were obtained, therefore, from an additional 42 hospital volunteers and 40 psychiatric patients. The volunteers consisted of 2 men and 40 women with a mean age of 33 years. The psychiatric patients consisted of 17 men and 23 women with a mean age of 34 years. The mean test-retest intervals were 1.0 for the volunteers and 2.2 for the patients. Thus, the standard-reverse correlations and reliabilities were obtained from similar samples and with equivalent test-retest intervals.

RESULTS

The product-moment correlation coefficients of true and false subscales with their reverse forms as well as the test-retest reliability coefficients are presented in Table 1. The significance of the difference between stand-

TABLE 1

COMPARISON OF TEST-RETEST COEFFICIENTS ON THE MMPI FOR SCALES IN STANDARD-REVERSE AND STANDARD-STANDARD FORMATS

MMPI scales	Psychiatric samples				Volunteer-college samples			
	True items		False items		True items		False items	
	A	B	A	B	C	D	C	D
L	a		.68	.71	a		.90	.88
F	.70	.77	.84	.88	.66	.76	.78	.79
K	a		.57	.72	a		.72	.87*
Hs	.78	.72	.79	.83	.70	.86*	.87	.85
D	.68	.84*	.58	.82*	.86	.88	.75	.75
Hy	.46	.76**	.39	.91**	.86	.74	.64	.75
Pd	.70	.78	.52	.59	.85	.83	.71	.80
Mf	.80	.80	.67	.89*	.82	.86	.85	.86
Pa	.72	.81	.50	.63	.79	.81	.45	.73*
Pt	.70	.68	.35	.40	.91	.85	.81	.70
Sc	.74	.72	.73	.63	.83	.76	.60	.67
Ma	.66	.83*	.63	.51	.67	.80	.80	.69
Mean	.69	.77	.61	.70	.79	.81	.74	.78

Note.—Samples A ($N = 54$) and C ($N = 35$) were given scales in standard-reverse format. Samples B ($N = 40$) and D ($N = 42$) were given scales in standard-standard format.

[a] *L* has no true items in its standard format, *K* has only one true item.

* $p < .05$, A versus B or C versus D, one-tailed test.

** $p < .01$, A versus B or C versus D, one-tailed test.

ard-reverse correlations and reliabilities was determined using the procedure given by McNemar (1949, p. 124). For the volunteer sample, the standard-reverse correlations are quite high, with a mean correlation of .79 for the true subscales and .74 for the false subscales. The standard-reverse correlations for the volunteers closely approximate the reliabilities; only 3 out of the 22 correlations are significantly different from their reliabilities.

For the patients, the standard-reverse correlations, while lower than for the volunteers, are still high. The mean correlation for the true subscales is .69 and for the false subscales .61. Six of the 22 correlations are significantly lower than their reliabilities, with the *Hy* scale showing the largest difference.

DISCUSSION

The magnitudes of the standard-reverse correlations are quite high, indicating that both patients and volunteers paid considerable attention to item content. The high correlations obtained are in marked contrast to the correlations found in item-reversal studies of the California F scale. For example, in the studies summarized by Christie et al. (1958) the correlations between regular F and reversed F ranged from −.66 to +.42.

The general correspondence of the standard-reverse correlations to the subscale reliabilities suggests that there is relatively little variance that could be attributed to acquiescence. This is particularly striking since the item reversals cannot be assumed to have perfectly reproduced the original content and therefore it is expected that the standard-reverse correlations would be lower than the reliabilities. While several scales may possibly reflect acquiescence variance, there is no consistency across the two samples as to which scales are vulnerable.

The present findings are in sharp contrast to the conclusions of Jackson and Messick that there is substantial acquiescence variance on the MMPI. Edwards and his students suggest that apparent acquiescence effects may be due to the confounding of indices of acquiescence with social desirability (Edward & Diers, 1962). Edwards (1961) has hypothesized that only items relatively neutral in social desirability will be vulnerable to acquiescence and Diers (1964) has provided some empirical evidence to support this hypothesis. Inspection of the desirability scale values for MMPI items presented by Messick and Jackson (1961b) indicates that about 17% of the items fall within the neutral range (scale values from 4.5 to 5.5 on a 9-point scale). Since relatively few MMPI items have neutral social desirability and if acquiescence operates primarily with neutral items, it would not be surprising to find minimal acquiescence effects.

Jackson and Messick (1962) agree that social desirability and acquiescence interact and concede that separately scoring true- and false-keyed subscales may have highlighted acquiescence effects. If the present item reversals are relatively successful in reproducing content, they also should reproduce equivalent desirability values. That is, a "true" reply to the standard item and a "false" reply to the reversed item would have about equal desirability values. With the item-reversal technique, therefore, the effect of acquiescence is independent of desirability (and content) in contrast with factor analytic technique in which these effects are confounded. In this study where acquiescence is not confounded with desirability, acquiescence is seen to contribute little variance to MMPI scales.

There is, as yet, no clear basis for deciding between the different conclusions reached by the factor analytic and item-reversal techniques. The data reported here, however, would seem to require that Jackson and Messick's conclusions with respect to acquiescence be considered tentative. Several other recent reports also suggest the need for caution. Siller and Chipman (1963) found that acquiescence accounted for little variance in a large battery of self-report measures. Dicken (1963) found that using acquiescence as a suppressor variable did not improve validity coefficients for the California Psychological Inventory. The importance of acquiescence for self-report inventories may have been overestimated.

REFERENCES

CHAPMAN, L. J., & CAMPBELL, D. T. Absence of acquiescence response set in the Taylor Manifest

Anxiety Scale. *Journal of Consulting Psychology,* 1959, **23**, 465–466.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the *F* scale irreversible? *Journal of Abnormal and Social Psychology,* 1958, **56**, 143–159.

DICKEN, C. Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement,* 1963, **23**, 699–720.

DIERS, CAROL J. Social desirability or acquiescence in response to personality items. *Journal of Consulting Psychology,* 1964, **28**, 71–77.

EDWARDS, A. E. Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *Journal of Abnormal and Social Psychology,* 1961, **63**, 351–359.

EDWARDS, A. E., & DIERS, CAROL J. Social desirability and the factorial interpretation of the MMPI. *Educational and Psychological Measurement,* 1962, **22**, 501–509.

JACKSON, D. N., & MESSICK, S. Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement,* 1961, **21**, 771–790.

JACKSON, D. N., & MESSICK, S. Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology,* 1962, **65**, 285–299.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

MESSICK, S., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin,* 1961, **58**, 299–304. (a)

MESSICK, S., & JACKSON, D. N. Desirability scale value and dispersions for MMPI items. *Psychological Reports,* 1961, **8**, 409–414. (b)

SILLER, J., & CHIPMAN, A. Response set paralysis; implications for measurement and control. *Journal of Consulting Psychology,* 1963, **27**, 432–438.

TAYLOR, JANET A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology,* 1953, **48**, 285–290.

(Received July 9, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 294–302

CONCEPTUAL ANALYSIS OF PSYCHOLOGICAL TEST SCORES AND OTHER DIAGNOSTIC VARIABLES[1]

HARRISON G. GOUGH

University of California, Berkeley

For the clinician, the validity of a test or assessment technique resides in the range and structural clarity of the information it provides him about the individual client with whom he works. 2 devices, for example, might be equal in forecasting a particular criterion, yet differ widely in their personological implicatons. Analysis of this problem permits specification of 3 levels of evaluation: primary, secondary, and tertiary. The conceptual model defined by these levels would appear to incorporate the kind of information which the diagnostician desires, and which indeed he must have if he is to function in an insightful and fully professional manner.

The purpose of this paper is to offer a point of view concerning the meaning of measurement in psychology. Attention is centered on the use or application of such measures, and what they tell the interpreter about the individual who has been tested. Significant prior discussions of the validity issue in testing and diagnosis have sought to classify tests according to the criteria employed in their construction and evaluation (cf. Cronbach & Meehl, 1955), and to specify ways in which discriminations may be sharpened (cf. Campbell & Fiske, 1959). These emphases are important, but nevertheless do not touch on all of the significant facets of meaning subsumed under the concept of validity. The intention here is to present a different perspective, one which stresses the implications of any scale or variable when it is brought to bear upon the analysis of the individual case.

From this perspective, the practitioner in testing seeks variables which permit individuated descriptions of the subject who has been tested, forecasts of what he will say or do, and characterizations of the way in which others will react to him. The greater the range of such information, and the more accurate its specification, the greater the value of the instrument which produced it.

An adequate theoretical position, following this, is one which points clearly to the kind of information which the practitioner must have if he is to employ his diagnostic tools in a responsible and professional manner. It should also provide help in structuring and clarifying this information, so that he may readily see what it is that is known, and what still needs to be known.

Two assumptions in this formulation are (*a*) that the purpose of the test is to assess and/or forecast significant nontest behavior, and (*b*) that the test is intended ultimately for interpretation and analysis of the individual case. A third principle is that the organization and application of this knowledge can only be realized through the endeavors of the trained, professional practitioner.

Our problem in defining criteria of validity and meaning is therefore to keep in mind the needs of the clinician who will interpret the tests, and the individual case which provides the fundamental context for their application. This task is not easy, and it is not expected that all readers will agree with the solution attempted in this paper. It may even be that each practitioner must solve these issues for himself, or try to solve them, and that no single proposal can win general acceptance. The only warrant for the present effort, if this is true, is that even in disagreement the reader may be stimulated to new considerations and perhaps to a more logical and clinically meaningful personal perspective.

THE CONCEPTUAL MODEL

Let us begin with a brief formulation of the conceptual model to be proposed, and after that move to more extended discussion of its components and to illustrative examples. This model is organized around three stages in the evaluative processes, that is, three foci of understanding which the interpreter must fully comprehend if he is to achieve an adequate conceptualization of a diagnostic variable. The first of these emphases may be designated the *primary evaluation*. The task here is to determine what criteria are principally relevant to the test, how well it predicts what it seeks to predict, measures what it purports to measure, or defines what it is intended to define. Most of what is said about test validity in textbooks and manuals may be classified under this first heading.

---

[1] A preliminary version of this paper was presented in a symposium on "Approaches to Integrated Concepts of Personality Theory and Assessment," New York Society for Projective Techniques, New York, June 6, 1964.

Secondary evaluation seeks to discover the psychological basis of measurement, to specify and clarify the meaning of that which is is measured. A scale may forecast some important nontest behavior, such as the likelihood of improvement in psychotherapy, and thus meet the requirements of our primary evaluation; but the clinician needs to know more than this, he must know what it is that the scale reflects that leads to this favorable potentiality. The task is to uncover and hence illuminate the underlying psychological dimensionality that is inherent in any test or measure possessing primary utility. When we return for specific consideration of this topic I shall attempt to enumerate explicit steps which can be taken in pursuit of this psychological (to be distinguished from psychometric) understanding.

Tertiary evaluation is perhaps more difficult to define than the other two concepts. It is concerned with the justification for developing a particular measure, or for calling attention to a measure. Part of this justification will come from the intrinsic significance of the primary aim of measurement, and part may come from the range of implications delineated in the secondary analysis. But additional and possibly even greater significance may come from the spectrum of life settings, beyond any envisaged under its primary validity, for which the technique has predictive and explanatory relevance.

An ability test might be developed with a primary goal of predicting success in school; on analysis, it is found that it does this satisfactorily, but perhaps no more so than many other such tests. Under secondary evaluation it is discovered that the key psychological variable which seems to be involved is the ability to reorganize and recombine perceptions and experiences, rather than the memory for facts and events which predominates in other devices. Our interest in this new test might now be heightening. But suppose that on the tertiary level we find, contrary to nearly all research with "ordinary" scholastic aptitude tests, that our new device appears to relate significantly to indices of creative and original endeavor, and that its forecasting efficiency becomes better and better as we move farther and farther away from the explicitly academic criteria in whose behalf it was initially constructed.

This kind of evidence is just what one seeks in the tertiary analysis; it is the kind of evidence which arouses the special interest of the clinician and which justifies his paying special attention to the instrument which possesses it.

PRIMARY EVALUATION

It might be useful to link the discussions of each of the three stages of analysis by offering observations of a single diagnostic variable under each heading. For this purpose I should like to utilize the So or "socialization" scale of the *California Psychological Inventory* (Gough, 1957). The CPI is a true-false objective inventory scaled for "folk concepts," that is, variables used for the description and analysis of personality in everyday life and in social interaction. It is theorized that such folk concepts, viewed as emergents from interpersonal behavior, have a kind of immediate meaningfulness and universal relevance which enhance their attractiveness as diagnostic concepts. Hopefully, diagnoses and forecasts of social behavior, if mediated by such concepts, will be more accurate and dependable than forecasts arrived at by way of other formulations.

The system of concepts defined by the 18 scales in the inventory should also be mentioned. Its main property is that it is an "open" system—it can change and grow by the addition or substitution of scales, and if need be can be reduced by the elimination of measures. Its purpose is to reflect social behavior, and in so doing to include a sufficient number of variables so that all major forms of such behavior can be forecast either by one scale or a combination or pattern of scales.

Within the full set of 18 variables there are certain natural clusterings, one of which refers to the domain of interpersonal values and intrapersonal controls. Three scales in particular bear on these issues: responsibility, socialization, and self-control. Although substantially intercorrelated, each is addressed to a different facet of the value complex: responsibility emphasizes the degree to which values and controls are conceptualized and understood; socialization emphasizes the degree to which they are internalized and made operational in the life of the individual; and self-control stresses the degree to which the individual approves of and espouses such regulatory dispositions.

Given these definitions, the primary validational task of the socialization scale is to locate individuals and groups along a continuum of asocial to social behavior, and to forecast the likelihood that any person will transgress whatever dividing line his own culture interposes between these two poles of the continuum. The phrase "his own culture" is used intentionally, because the folk concept theoretical basis of the inventory requires that its validity be demonstrated in other cultures than the one in which the sale is developed.

The So scale was first introduced in 1952 (Gough & Peterson, 1952). In 1960 a report was published (Gough, 1960b) which surveyed validational evidence obtained in the preceding 8-year period. For males, 25 samples involving 10,296 cases were considered, ranging from nominated "best citizens," through occupational samples of varying kinds, disciplinary problems, and county jail inmates, to incarcerated delinquents and felons. The biserial correlation for the dichotomy of more- versus less-socialized was +.73. For females, 16 samples totaling 10,560 subjects were studied, covering the same continuum of socialization; the biserial correlation here was +.78.

Primary validation evidence for classification of individuals along the full extent of the socialization continuum would appear adequate. The next question pertains to differentiation within zones or regions of the continuum. An interesting finding with respect to this question comes from the work of Vincent (1961) on unmarried mothers; for a sample of 232 women the biserial correlation between the So scale and the dichotomy of one illegitimate pregnancy versus two or more was +.83. Highly significant differences between recidivists and first offenders were found in studies of reformatory inmates (Donald, 1955) and juvenile offenders (Peterson, Quay, & Anderson, 1959).

Cross-cultural validation has been equally encouraging. For a sample of 203 institutionalized delinquents in India, tested in Hindi and Punjabi (Gough & Sandhu, 1964), the So scale correlated +.73 with court-assigned classifications of the severity of the offense. Just now we are reviewing data from Austria, Costa Rica, France, Germany, Italy, Japan, Puerto Rico, South Africa, and Switzerland, comparing delinquent versus nondelinquent samples of males and females. Nine countries are involved in this survey, and translations of the So scale into six languages: Afrikaans, French, German, Italian, Japanese, and Spanish. In every comparison the So scale has differentiated significantly between delinquents and nondelinquents; and, considering all samples, in no instances does the average for any sample of delinquents (no matter from which country or place) equal or exceed the lowest average score observed among the samples of nondelinquents.

There is also evidence from a study in progress by Ernest Wenk and the author on the prediction of parole outcomes. For a sample of 295 parolees followed for 3 years on parole it was found that 165 men were successful and 130 were violators. The So scale means for these two samples were 26.84 and 24.77; this difference, although not large, yields a *t* ratio of 3.33, which is significant beyond the .01 level of confidence.

The evidence just reviewed for the So scale is the kind of evidence we customarily expect on validity, and it is the kind of evidence which we must establish in our primary evaluation. The evaluation is primary because it comes first, and because there is no reason to follow a measure on into the realms of secondary and tertiary evaluation unless we are reasonably satisfied that its primary utility is established.

SECONDARY EVALUATION

We are now ready to undertake a secondary evaluation of the So scale. Our aim here, to repeat, is to clarify the basis of measurement, that is, to determine what it is that the scale reveals about a person beyond the fact that he will probably behave in a more- or less-socialized manner. The four steps enumerated below are offered as aids in the accomplishing of this conceptual analysis.

> Step 1. Review of the development of the measure, the procedures and samples used in its construction, its theoretical presuppositions and bases.
>
> Step 2. Analysis of the components of the measure, its items, stimulus materials, and content.
>
> Step 3. Determination of the relationships between the measure and (*a*) other measures already known and conceptualized, and (*b*) variables of self-evident importance such as sex, age, status, etc.
>
> Step 4. Specification of the characterological and personological dispositions of individuals who obtain scores defined by the measure itself as diagnostically significant.

A primary evaluation will already have provided some of the information envisaged under Step 1, but other matters await consideration. Perhaps the most important of these is the conceptual basis of the measure. The So scale was initially conceived in the context of role theory. The less-socialized person was hypothesized to be less skillful in sensing and interpreting subtle and covert cues in social interaction, and hence less likely to evolve dependable and veridical internalized systems of control. Indirect evidence of such role-taking disability would be found in rule-breaking and rule-violating behavior, and direct evidence would be found in tests of the accuracy of role perspectives and interpersonal diagnosis.

In the construction of So, samples ranging from seriously delinquent through moderately delinquent to conventionally socialized were used, and item studies undertaken so as to identify those possessing validity for this continuum. The assumption here, the reader will recognize, is that one's manifest location along the socialization continuum will on the average provide a valid estimate of one's role-taking proficiency. Whether or not this assumption is true or partly true, a more direct test of the hypothesis is clearly needed.

To date, the most thorough check of this sort is found in a paper in 1957 by Reed and Cuadra. They studied 204 student nurses in a neuropsychiatric hospital. Each nurse described herself on the Gough Adjective Check List (ACL) (Gough, 1960a; Gough & Heilbrun, 1965) and then the other three nurses in a four-member group to which each was assigned. Next, each nurse attempted to guess how she would be described by her group. A point was earned for a predicted adjective if two of the three peers had, in fact, checked this word, and a total score was defined as the sum of these points. This score is therefore representative of the accuracy with which an individual senses the reactions of others to her, that is, of the degree to which she can take the role of the other and look upon self as an object. The correlation between the So scale and this social sensitivity index was +.41.

A related task would be to ask the individual to guess how another person would describe himself. In an unpublished study at San Quentin Prison in California, Bela Baker had inmates guess how their cellmates had answered the ACL; the accuracy of these estimates was significantly correlated with the inmates' own scores on the So scale.

Recapitulating briefly the information under Step 1, we can say that the scale may be conceptualized from the role-taking perspective and that there is both direct and indirect evidence for its validity from this perspective.

Step 2 takes us into the content of the scale. Here we find that its 54 items are about equally distributed over two kinds of content. The first of these more or less directly embodies role-taking ideas, and may be instanced by these items: "Before I do something I try to consider how my friends will react to it," "I often think about how I look and what impression I am making upon others," and "I find it easy to 'drop' or 'break with' a friend."

The other group of items has more immediate relevance to deviant or rule-breaking behavior, this being justified by the aforementioned hypothesis that asocial deviation is itself an indicator of the role-taking disability. Examples of this second type of item are: "I have often gone against my parents' wishes," and "If the pay was right I would like to travel with a circus or carnival."

The task for the practitioner in this second step is to internalize the content of the scale, and by virtue of this intimate familiarity to enhance the insightfulness of the psychodynamic formulation which is being evolved. The clinician who has studied and restudied the Rorschach inkblots and the cards of the Thematic Apperception Test need not be told that this second step, the personal, intense, apperceptive, and empathic perusal of the stimulus materials included in the scale or instrument is one of compelling importance and significance. It is unfortunate indeed that this step, invariably taken by projective testers, is only occasionally taken by psychologists in the ranks of objective or structured testing.

Step 3 concerns the relationships of the measure to other measures and to major categories of interpersonal variation. Perhaps the first question that should be asked of any test or scale being studied is "does it show a sex difference?" and the second "is it correlated with intelligence?" Among nondelinquent samples, both in the United States and elsewhere, the So scale does reveal a consistent sex difference of about half a standard deviation, with women scoring higher. One is tempted to say "as would be expected," as many would assert that women are in fact more law-abiding, more sensitive and perceptive, and more highly refined and elaborated in their role-taking. However, when attention is directed to delinquent and institutionalized samples, women tend to score a bit lower than their male counterparts.

Role-taking skill, if that is the underlying variable in the So scale, is probably correlated with intellect, but one would nonetheless hope that a personological measure could be defined which would be free of this particular component. The evidence seems to suggest that So is not correlated with intellectual ability. In seven or eight samples of fairly good size, varying with respect to age, educational level, etc., the correlations fall within a band of $-.14$ to $+.10$, with a median of close to .00.

Although So is unrelated to intelligence, it is related to operant conditioning under social reinforcement. Sarbin, Allen, and Rutherford (1965) compared 30 chronically delinquent boys with 30 nondelinquents, matched on age, intelligence, and occupation of parents, seeking evidence on the social responsiveness of delinquents to verbal conditioning. On two conditioning trials the delinquents did not differ significantly from the controls, but when higher-scoring delinquents on the So scale were compared with lower-scoring delinquents, significant differences in conditioning (favoring the higher scorers) were obtained; the same findings were observed among the controls, where those with higher scores on So conditioned more readily in the social learning task than those with lower scores.

What about other variables? Age norms would be of great interest, but unfortunately are not available. There is a hint, however, that the So scale rises more rapidly in adolescence and approaches the adult level earlier than do scores on the CPI scales for social presence, dominance, and self-acceptance.

Social status is another variable of interest. In several high schools the So scale yielded a median coefficient of $+.11$ with the Home Index (Gough, 1949), a measure of the socioeconomic status of the home and family background. Race differences have also been investigated, by Donald (1955) and Peterson et al. (1959) in the studies previously mentioned; in both instances Negro and white delinquents were *not* differentiated on the So scale.

Social desirability as a basis for response to the So scale might also be a source of worry. In a sample of 295 males, the correlation between So and Edwards' 39-item Social Desirability scale (Edwards, 1957) was $+.21$. In a sample of 152 males So correlated $+.01$ with Barron's scale for ego strength (Barron, 1953), $+.11$ with Welsh's anxiety index (Welsh, 1952), and $+.22$ and $-.23$ with the K and Psychopathic Deviate scales of the MMPI (Hathaway & McKinley, 1943).

A great deal of information of this sort is offered in the CPI *Manual* (Gough, 1957), and may be consulted there by anyone wish-ing to push further into this third category. What has been presented is believed to be representative, and would seem to justify the conclusion that the So scale is a relatively independent variable, not contaminated by or unduly influenced by such factors as social desirability, status, anxiety, race, intellectual ability, or ego strength. Hence, it is not to be "explained" or "explained away" by such factors, but must be dealt with in its own right.

This leads us to Step 4, and a consideration of the immediate and direct personological implications of the scale. Although the previous steps are of interest to the clinician, and in fact vital if he wishes to reach a sophisticated and professional level of insight concerning any measure, it is the fourth step which comes closest to a direct delineation of the psychological meaning of a variable.

There are different ways of developing the information called for by this fourth step. One is the time-tested method of the individual clinician, who notes over the years of his practice what it is that characterizes patients scoring high or low on a particular test index—their defenses, their strengths, attitudes, self-conceptions, and stylistic predilections. The problem in this approach is that, even if valid, it takes years of patient accumulation of evidence before any sort of pattern begins to emerge. The need is for a faster and more efficient technique which retains the essential validity of the clinician's observations but shortens the time necessary to obtain them and to process them.

The personality assessment method, as practiced at the Institute of Personality Assessment and Research in Berkeley and at other centers, is exceedingly well adapted to do just this. Subjects come to the Institute in small numbers, and are studied intensively for 3 or 4 days by a panel of from 5 to 10 psychologists. Trait ratings, interviewers' formulations, *Q*-sort appraisals, and adjective check list descriptions are systematically gathered on all assessees. These sources of observational and diagnostic data may then be related back to the test scores of the assessees. The process is perhaps most clearly illustrated by ACL analyses, and only ACL data will therefore be drawn on in the discussion to follow.

Suppose 10 observers in assessment each complete a 300-word Gough Adjective Check List on an assessee. Any trait or quality which is above threshold in the eyes of the observer is checked, and any which is not is left blank. A convenient composite can be derived from these 10 individual descriptions simply by counting the number of times a word is checked, and treating that total as a score. Thus, if an assessee Adams is checked by all 10 observers as "alert" his score on this interpersonal quality or trait is 10; if 2 observers check him as "blustery" his score is 2; and if no one checks him as "charming" his score is zero. Adams will in this way be assigned 300 scores, 1 for each adjective.

If a sample of assessees has been described by the same panel of observers, these adjectival scores may be treated correlationally. A test variable, for example, the So scale, may be correlated with each of the 300 words in this sample of assessees. Adjectives showing significant and positive correlations are those which tend to be used to characterize high scorers on the scale, and hence afford a conceptual starting point for a personological sketch of the high scorer. Likewise with adjectives showing significant and negative correlations with the scale: those are the words that are in fact differentially used to characterize low scorers and they afford a valid starting point for a personological formulation of the low scorer. I believe that this procedure is intrinsically valid, and that it offers an efficient and powerful method for the psychological analysis of any variable that can be dichotomized or quantified. One thinks first of scales or single scores, such as for So, but the method is equally applicable to patterns, configurations, regression equations, or any other method of classifying, ranking, or calibrating individuals.

This method of analysis has been applied five or six times to the So scale. I should like to draw on four of these analyses, one based on a sample of 295 adult males, a second on a sample of 80 university graduate students, a third on a sample of 51 college seniors, and the fourth a sample of 100 military personnel studied by Reed and Cuadra. The first two samples are male, the third female, and the fourth includes both males and females; in three samples the observations were contributed by psychologists who had studied the assessees, and in the last the descriptions were furnished by peers. The purpose in using all four samples is to overcome any limitations which might attach to a particular setting, use of professional versus peer evaluations, and to sex differences.

From these four independent analyses, those adjectives correlating in the same direction with the So scale in all four instances, and at a statistically significant level in at least two are presented below:

A. Adjectives used significantly more often to describe high scorers on the So scale.

calm
considerate
conventional
cooperative
helpful
moderate
modest
obliging
patient
peaceable
steady
trusting
unassuming

B. Adjectives used significantly more often to describe low scorers on the So scale.

affected
arrogant
conceited
cynical
defensive
disorderly
dissatisfied
headstrong
impatient
impulsive
irritable
rebellious
restless
self-centered
wary

A comment should be interjected here. Some psychologists, perhaps not given to thinking clinically, seem to be more bothered than helped by such a list of attributes. Such persons, one fears, are beyond the reach of the theoretical position being advocated in this paper. These adjectival clusters represent only the elements of clinical description, that is to say, protocol observations. From this starting point it is the interpreter's task to evolve an insightful diagnostic portrait of the high and low scorer on the scale, to render this evidence into an integrated formulation relevant for practice and understanding. The achievement of such a formulation requires all of the creativity that the interpreter can muster. But whatever its difficulty, the task must be attempted, for the elaboration of a psychological dynamic of the test variable is at the very heart of the conceptual analysis which is being illustrated.

TERTIARY EVALUATION

We come now to the third perspective, that of tertiary analysis. Is there any reason for paying attention to a scale or variable if one is not interested in its primary validational focus, that is, with respect to the So scale, in the problem of asocial versus socialized behavior?

Let us say that we grant that So can forecast asocial behavior with a surprisingly high degree of accuracy, and that it rests on a rather interesting theoretical and personological basis. Let us grant further that it is free of nuisance correlations with such variables as response sets, intellectual ability, and socioeconomic status. But if we are not working in a prison or juvenile hall, and if we are not concerned with identifying asocial dispositions, is there any other reason for studying this scale and for learning how to utilize it clinically? These are the questions met in the tertiary evaluation, questions which a systematic theory of test meaning must attempt to answer. We need to show, in other words, that the variable is in fact of significance in situations other than those encompassed by its domain of primary relevance.

For the So scale, three examples of such tertiary significance may be offered. The first comes from work on differential achievement among persons of unusually high intellectual talent. In 1955 a nationwide sample of high school and college students was surveyed (Gough, 1955), searching for correlates of differential academic achievement among students in the top 5% to 10% of the aptitude distribution. The one variable from the CPI which stood out above all others was the So scale: its correlations were significant for both sexes and at both educational levels. Superior achievement, as evidenced by grades, was associated with higher So scores. Note, in considering these results, that the So scale shows little or no correlation with grades if an unselected sample is studied; the CPI contains other scales, such as the two achievement indices, which are much more relevant under these circumstances.

These findings led to the theoretical formulation that consistent use of high-level talent in the academic setting is more a matter of socialization and the cathexis of approved goals and objectives than of achievement motivation per se, or of ambition as ordinarily defined and appraised, or of negative spurs such as anxiety and self-doubt.

This early study was followed by a much larger and more definitive analysis by Holland in 1959, in which freshman grades in college were forecast for a sample of 1,321 National Merit Scholarship corporation finalists. Holland found that the So scale gave valid forecasts of achievement as reflected in grades, and he also observed that the correlations for the So scale were two and three times as high as those for the Scholastic Aptitude Tests given after the finalists had already been selected.

Here, then, is one domain of functioning—academic achievement among the intellectually gifted—in which the So scale has a special significance and in which it can make a contribution which up to this time has been matched by no other variable.

A second example of a tertiary implication comes from studies of college graduation. It is known that only 40% to 50% of all students who begin college go on to graduation. This is a serious problem to educators, behavioral scientists, and everyone else. Would the So scale, administered at entry into college, be of any diagnostic value in identifying potential dropouts? The author is just finishing a large study of 3,242 students from seven classes in six colleges, in which CPI scales and Scholastic Aptitude Test scores are being compared on their forecasting efficiency. All test scores were obtained at the time of admission to college, and criteria on graduation versus dropping out were obtained 4 or more years later.

The aptitude test scores do not differeniate very well, contrary to the expectations of most psychologists. Such tests do predict GPA in college, but evidence is mounting that they bear less relationship to criteria of either rate of progress or survival (cf. Glimp & Whitla, 1964). The So scale, on the other hand, appears to be a useful predictor of graduation. Potential graduates achieve higher So scores at admission, and potential dropouts attain lower scores.

But is there an important social criterion on which persons with lower So scores excel? Certainly we should not expect or wish that in any and all situations the more conventional, more stringently self-regulated, and more interpersonally adaptable individual would do best. One such happy exception is found in the studies of creativity. Barron (1961) in his comparison of creative and journeyman writers found the former to score significantly lower on the So scale. MacKinnon (1961) in his study with Hall of 124 American architects obtained the same findings. The most creative architects scored lowest on the So scale, the least creative scored highest, and the intermediate group on creativity occupied an intermediate point on the So distribution.

CONCLUSION

What is the purpose of the method of analysis presented in this paper? Its principal goal is to specify the kind of information and comprehension needed by the practitioner of testing. The position taken is that the user of any test must become intimately and fully familiar with all of the kinds of evidence sketched here for the So scale if he is to apply the instrument in a responsible and professional manner. Analysis of the conceptual problem into primary, secondary, and tertiary components may help him to keep straight what he must do, and may indicate the kinds of information he must discover and then assimilate. Out of all this he seeks to determine what it is that he can say about the individual who has been tested. The hope is to achieve a true, profound, and individual portrait of the person being appraised.

But alas, there is no rest for the weary, and once this is done for a single scale or component of a multivariate instrument it must be done for all other components. It is a difficult task for anyone to master the meanings and implications of the variety of responses subjects make to shading on the Rorschach. But even if this mastery is achieved, is the clinician ready to interpret all other responses? No indeed, for he must come to comprehend each component in this fully dynamic and articulated fashion.

And even here we are only at the beginning and not the end of the training of the diagnostician, for once components are understood the more challenging and complex issues which arise in the interpretation of configurations and patterns must be resolved. This paper has dealt almost entirely with what one must do to arrive at an adequate comprehension of one component of one test; certainly it is evident that this is only a first step toward the kind of organized and insightful interpretation of profiles and test batteries which is the ultimate goal of the clinician in diagnosis. This inspiring and ultimate edifice of test usage, clearly, can only be constructed from building blocks of the highest quality; our need, therefore, is to delineate principles which can at each level contribute significantly to the generation of valid, dependable, and diagnostically relevant information.

REFERENCES

BARRON, F. An ego-strength scale which predicts response to psychotherapy. *Journal of Consulting Psychology*, 1953, **17**, 327–333.

BARRON, F. Creative vision and expression in writing and painting. In, *Proceedings of the Tahoe Conference on the Creative Person*. Part II. Berkeley, Calif.: Liberal Arts Department, University Extension, University of California, 1961. Pp. 1–19.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the mulitrait-multimethod matrix. *Psychological Bulletin*, 1959, **56**, 81–105.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, **52**, 281–302.

DONALD, E. P. Personality scale analysis of new admission to a reformatory. Unpublished master's thesis, Ohio State University, 1955.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

GLIMP, F. L., & WHITLA, D. K. Admissions and performance in college. *Harvard Alumni Bulletin*, 1964, **66**(7), 304–309.

GOUGH, H. G. A short social status inventory. *Journal of Educational Psychology*, 1949, **40**, 52–56.

GOUGH, H. G. Factors related to differential achievement among gifted persons. In P. A. Witty (Chm.), The gifted child. Symposium presented at the American Psychological Association, San Francisco, September 1955.

GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consulting Psychologists Press, 1957.

GOUGH, H. G. The Adjective Check List as a personality assessment research technique. *Psychological Reports*, 1960, **6**, 107–122. (a)

GOUGH, H. G. Theory and measurement of socialization. *Journal of Consulting Psychology*, 1960, **24**, 23–30. (b)

GOUGH, H. G., & HEILBRUN, A. B., JR. *Manual for the Adjective Check List*. Palo Alto: Consulting Psychologists Press, 1965.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology*, 1952, **16**, 207–212.

GOUGH, H. G., & SANDHU, H. S. Validation of the CPI Socialization scale in India. *Journal of Abnormal and Social Psychology*, 1964, **68**, 544–547.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1943.

HOLLAND, J. L. The prediction of college grades from the California Psychological Inventory and the Scholastic Aptitude Test. *Journal of Educational Psychology*, 1959, **50**, 135–142.

MACKINNON, D. W. Creativity in architects. In, *Proceedings of the Tahoe Conference on the Creative Person*. Part V. Berkeley, Calif.: Liberal Arts Department, University Extension, University of California, 1961. Pp. 1–24.

PETERSON, D. R., QUAY, H. C., & ANDERSON, A. C. Extending the construct validity of a socialization scale. *Journal of Consulting Psychology*, 1959, **23**, 182.

REED, C. F., & CUADRA, C. A. The role-taking hypothesis in delinquency. *Journal of Consulting Psychology*, 1957, **21**, 386–390.

SARBIN, T. R., ALLEN, V. L., & RUTHERFORD, E. E. Social reinforcement, socialization, and chronic delinquency. *British Journal of Social and Clinical Psychology*, 1965, **4**, 73–78.

VINCENT, C. *Unmarried mothers*. New York: Free Press of Glencoe, 1961.

WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *Journal of Consulting Psychology*, 1952, **16**, 65–72.

(Received July 27, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 303–306

TEMPORAL EFFECTS OF LSD-25 AND EPINEPHRINE ON VERBAL BEHAVIOR[1]

GILBERT HONIGFELD

Veterans Administration Central Neuropsychiatric Research Laboratory, Perry Point, Maryland

Cloze analysis was used to evaluate the temporal effects of LSD and epinephrine on the understandability of verbal communication. 10 normal male Ss recorded spontaneous speech samples before, and at 8 fixed intervals after injection of each of the drugs. Order of drug administration was varied randomly. Typescripts of these speech samples with every 5th word deleted were completed by 13 normal adults and scored according to the number of agreements with the original sources. The 2 drugs acted differentially ($p < .001$), and the expected interaction of drug and time effects was obtained ($p < .05$). Compared with predrug levels, LSD impaired understandability, while epinephrine enhanced understandability. Peak effects for both drugs were noted about 2 hours after injection.

Studies of the psychological effects of drugs in normal human subjects are usually concerned with such basic phenomena as perception or conditioning (Eysenck, 1963; Uhr, 1960). Effects of drugs on more complex cognitive functions, such as language, have been much less adequately studied. A feature of man's relationship with his environment which is responsive to the effects of drugs is the ability to communicate, or to be understood by others. As used here, understandability refers not to the adequacy with which speech sounds are produced, although drugs may affect this mechanical aspect of speech (Starkweather & Hargreaves, 1964), but rather to the extent to which verbal content is communicated. That is, this report will be concerned with the effects of drugs on verbal behavior, not vocal behavior. Previous studies of verbal behavior have applied the technique of Cloze analysis to evaluate the effects of anxiety (Platz & Honigfeld, in press), high drive levels (Osgood & Walker, 1959), authoritarianism (Honigfeld, Platz, & Gillis, 1964), psychopathology (Honigfeld, 1963b), psychotherapeutic interaction (Feldstein & Jaffe, 1963), and drugs (Honigfeld, 1963a; Salzinger, Pisoni, Feldman, & Bacon, 1961).

Cloze analysis is a method in which verbal material is mutilated by the deletion of every fifth (or n^{th}) word, with subjects being instructed to fill in the blanks as well as they can (Taylor, 1956). Expressed as the proportion of agreements with the source, the resulting index may be thought of as expressing the "overall correspondence . . . between the language systems of different individuals [Osgood, 1959, p. 80]." In other words, Cloze analysis provides an index of understandability.

A basic aspect of psychopharmacology concerns the temporal effects of acute drug administration. Most studies of the behavioral effects of drugs fail to present this kind of information. Postdrug testing is usually done at one fixed interval, although the appearance and magnitude of effects depend in part on the suitability of the interval selected for postdrug testing. In all of Uhr and Miller's (1960) encyclopedic work, there is not a single time-response curve showing the behavioral effects in man of any psychoactive drug. Yet, in the opening chapter of that same volume, Russell (1960) pointed out that "Because of this growth and decay of drug effectiveness, standardized measures of behavior taken at different times after ad-

---

[1] This research was supported in part by Grant No. MH-05144 from the National Institute of Mental Health. Thanks are due to Leo E. Hollister, Veterans Administration Hospital, Palo Alto, California, in whose laboratory these data were collected, and to Arthur Platz, Veterans Administration Hospital, Perry Point, Maryland, who provided valuable advice and encouragement. The assistance of the National Clearinghouse for Mental Health Information, National Institute of Mental Health, is gratefully acknowledged.

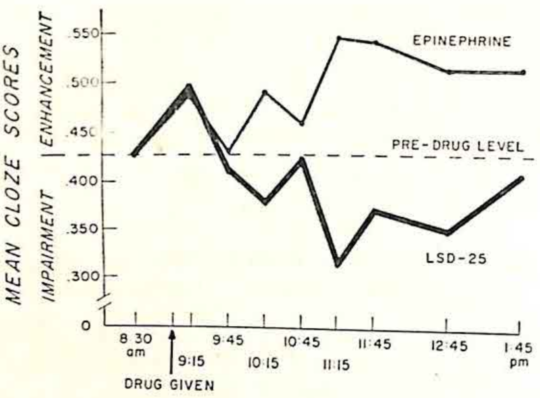


FIG. 1. Temporal effects of LSD-25 and epinephrine on understandability of spontaneous speech.

ministration of a drug may provide quite different results [p. 34]."

It was the purpose of this study to provide some comparative data on time-response relationships for LSD and epinephrine. Both these drugs may be expected to affect psychomotor and cognitive functioning (Abramson, Jarvik, & Hirsch, 1955; Aronson, Watermann, & Klee, 1962; Hirsch, Jarvik, & Abramson, 1956; Hoffer, Osmond, & Smythies, 1954; Kornetsky, Humphries, & Evarts, 1957; Krus & Wapner, 1962; Levine, Abramson, Kaufman, & Markham, 1955; Pollin & Goldin, 1961; Silverstein & Klee, 1958; Thuring, 1960). However, the effects of LSD and epinephrine on understandability had not been systematically examined previously. In a pilot study (Honigfeld, 1963a) another hallucinogenic agent, psilocybin, exerted maximum impairment of understandability between 1½ and 2½ hours after drug ingestion. However, that study was primarily of methodologic interest since it was the first such application of Cloze analysis to appear in the professional literature. In the present investigation, the same methods were extended in a study of the temporal effects of LSD and epinephrine.

METHOD

Ten adult male volunteer subjects in good physical and emotional health were studied individually. Tape recordings of each subject's speech were made in accordance with these instructions:

> In order to get a complete record of these experimental sessions, we'd like you to describe your experiences verbally.
>
> At each of the following times, you are to tape record your impressions of the experimental situation and your reactions to it:
>
> 8:30 A.M.
> 9:15 A.M.
> 9:45 A.M.
> 10:15 A.M.
> 10:45 A.M.
> 11:15 A.M.
> 11:45 A.M.
> 12:45 P.M.
> 1:45 P.M.
>
> Merely press the "Record" button, and speak into the microphone for about 5 minutes each time. Please announce the correct time whenever you press the "Record" button. When you're finished, press the "Stop" button. Feel free to speak into the microphone at *any time* during these sessions, but be certain that you do record your feelings and impressions at each of the 9 intervals already noted. Try to be as introspective as possible, going into as much detail as you feel is necessary.

At 9:00 A.M., subjects received a subcutaneous injection of either epinephrine (3.0 mcg/K) or LSD-25 (2.0 mcg/K), with local massage to hasten absorption. Testing sessions were spaced approximately 1 week apart, with half the subjects receiving LSD first, and half receiving epinephrine first.

Typescripts of each subject's 9 speech samples were prepared for Cloze analysis by the deletion of every fifth word. Thirteen evening college students comprised the panel of normal language users who completed the Cloze passages. Since each of the 10 drug subjects gave 9 speech samples under both experimental conditions, there was a total of 180 passages to complete. To make this task manageable, each mutilated passage was completed by 1 of the 13 student panelists, according to a random assignment procedure. In this way each panelist completed a total of 13 or 14 mutilated passages. Therefore, at each time period under either drug condition 10 student panelists (randomly picked out of the total 13) completed 10 mutilated passages. The data presented in Figure 1 represent the mean scores of 10 panelists, each completing one passage from a different speaker.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF CLOZE SCORES

Source	*df*	*MS*	*F*
Drugs (A)	1	384,662	24.45**
Time (B)	8	10,464	ns
Subjects (C)	9	15,400	ns
A × B	8	35,866	2.39*
A × C	9	21,450	ns
B × C	72	13,107	ns
A × B × C	72	15,019	ns
Total	179		

* $p < .05$, error term = A × B × C interaction (72 *df*).
** $p < .001$, error term = pooled A × B × C + A × C interaction (81 *df*).

RESULTS AND DISCUSSION

The analysis of variance model applied was for an experimental procedure replicated on each of several individuals (Walker & Lev, 1953). The design involved 10 Subjects × 2 Drugs × 9 Time Periods. It was expected that a significant main effect for drugs would emerge, and a significant interaction between drugs and time was also anticipated. The results reported in Table 1 support both hypotheses. Significant *F*s were found both for drugs ($p < .001$), and for the Drug × Time interaction ($p < .05$).

It is clear that these two drug conditions exerted differential effects on the intelligibility of spontaneous speech. Further, the emergence of these effects was related to the temporal action of the drugs themselves (see Figure 1). The predrug and first postdrug Cloze score means were identical under both LSD and epinephrine conditions. Within an hour after the drug injection, the trend that was established over the remaining period of observation began to emerge—under the epinephrine condition understandability was enhanced, while under LSD understandability was impaired. Peak effects occurred about 2 hours after drug administration, with a gradual return toward baseline levels. The LSD-epinephrine mean differences were statistically significant (*t* tests between correlated means) at 11:15 ($p < .05$), 11:45 ($p < .01$), and 12:45 ($p < .01$). The observation period was not extended long enough in time to yield a complete return to predrug levels, although the final difference in mean scores at 1:45 was not statistically significant.

REFERENCES

Abramson, H. A., Jarvik, M. E., & Hirsch, M. W. Lysergic acid diethylamide (LSD-25) VII. Effect upon two measures of motor performance. *Journal of Psychology,* 1955, 39, 455–464.

Aronson, H., Watermann, C. E., & Klee, G. D. The effect of D-lysergic acid diethylamide (LSD-25) on learning and retention. *Journal of Clinical and Experimental Psychopathology and Quarterly Review of Psychiatry and Neurology,* 1962, 23, 17–23.

Eysenck, H. J. Personality and drug effects. In H. J. Eysenck (Ed.), *Experiments with drugs.* New York: Macmillan, 1963.

Feldstein, S., & Jaffe, J. Language predictability as a function of psychotherapeutic interaction. *Journal of Consulting Psychology,* 1963, **27,** 123–126.

Hirsch, M. W., Jarvik, M. E., & Abramson, H. A. Lysergic acid diethylamide (LSD-25) XVIII. Effects of LSD-25 and six related drugs upon handwriting. *Journal of Psychology,* 1956, 41, 11–22.

Hoffer, A., Osmond, H., & Smythies, J. Schizophrenia: A new approach. II. Result of a year's research. *Journal of Mental Science,* 1954, **100,** 29–45.

Honigfeld, G. Effect of an hallucinogenic agent on verbal behavior. *Psychological Reports,* 1963, **13,** 383–385. (a)

Honigfeld, G. The ability of schizophrenics to understand normal, psychotic and pseudo-psychotic speech. *Diseases of the Nervous System,* 1963, **24,** 692–694. (b)

Honigfeld, G., Platz, A., & Gillis, R. Verbal style and personality: Authoritarianism. *Journal of Communication,* 1964, **14,** 215–218.

Kornetsky, C., Humphries, Ogretta, & Evarts, E. V. Comparison of psychological effects of certain centrally acting drugs in man. *American Medical Association Archives of Neurology and Psychiatry,* 1957, 77, 318–324.

Krus, D. M., & Wapner, S. Effect of LSD on pace of performing a variety of tasks. *Perceptual and Motor Skills,* 1962, **14,** 255–259.

Levine, A., Abramson, H. A., Kaufman, M. R., & Markham, S. Lysergic acid diethylamide (LSD-25) XVI. The effect on intellectual functioning as measured by the Wechsler-Bellevue Intelligence Scale. *Journal of Psychology,* 1955, **40,** 385–395.

Osgood, C. E. The representational model and relevant research methods. In I. deS. Pool (Ed.), *Trends in content analysis.* Urbana, Ill.: Univer. Illinois Press, 1959. Pp. 33–88.

Osgood, C. E., & Walker, Evelyn G. Motivation and language behavior: A content analysis of suicide notes. *Journal of Abnormal and Social Psychology,* 1959, **59,** 58–67.

Platz, A., & Honigfeld, G. Some effects of anxiety on the intelligibility of verbal communication in psychotherapy. *Journal of Personality and Social Psychology,* 1965, **2,** 122–125.

Pollin, W., & Goldin, S. The physiological and psychological effects of intravenously administered epinephrine and its metabolism in normal and schizophrenic men—II. Psychiatric observations. *Journal of Psychiatric Research,* 1961, **1,** 50–67.

Russell, R. W. Drugs as tools in behavioral research. In L. Uhr & J. G. Miller (Eds.), *Drugs and behavior.* New York: Wiley, 1960. Pp. 19–40.

Salzinger, K., Pisoni, Stephanie, Feldman, P. S., & Bacon, Pauline M. The effect of drugs on verbal behavior. In, Control of Verbal Behavior. Symposium presented at American Association for the Advancement of Science, Denver, December 1961.

Silverstein, A. B., & Klee, G. D. Effects of lysergic acid diethylamide (LSD-25) on intellectual functions. *American Medical Association Archives of Neurology and Psychiatry,* 1958, **80,** 345–350.

STARKWEATHER, J. A., & HARGREAVES, W. A. The influence of sodium pentobarbital on vocal behavior. *Journal of Abnormal and Social Psychology*, 1964, **69**, 123–126.

TAYLOR, W. L. Recent developments in the use of "Cloze procedure." *Journalism Quarterly*, 1956, **33**, 42–46.

THURING, J. P. The influence of LSD on the handwriting pressure curve. *Advances in Psychosomatic Medicine*, 1960, **1**, 212–216.

UHR, L. Objectively measured behavioral effects of psychoactive drugs. In L. Uhr & J. G. Miller, (Eds.), *Drugs and behavior*. New York: Wiley, 1960. Pp. 610–633.

UHR, L., & MILLER, J. G. *Drugs and behavior*. New York: Wiley, 1960.

WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953. Pp. 363–372.

(Received August 4, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 307–311

RELATION BETWEEN BARRIER SCORES AND AGGRESSIVE BEHAVIOR

EDWIN I. MEGARGEE[1]

University of Texas

A study by Fisher and Cleveland (1958) indicated that high barrier Ss express anger outwardly, yet other investigations have indicated low barrier Ss are more impulsive and aggressive. A study of barrier scores in a sample of 75 juvenile delinquents indicated that the delinquents had barrier scores significantly lower than nondelinquent adolescents; moreover the more seriously delinquent Ss had barrier scores significantly lower than the less seriously delinquent Ss. Correlational data also indicated some association between low barrier scores and aggressiveness. The discussion focused on the adaptiveness of aggression in various situations and on the interpretation of the barrier score as an index of ego identity as opposed to ego diffusion.

In their original presentation of the inkblot barrier score as a measure of the definiteness of body boundary perception, Fisher and Cleveland (1958) noted an apparent inconsistency in the relation of barrier scores to impulsivity and aggressiveness. In a study of 69 Harvard undergraduates, they found that those who had high barrier scores were more likely to express their anger outwardly after undergoing a humiliating and frustrating experience, while those who had low barrier scores were more inclined to direct their anger inward in a self-punitive fashion.

On the other hand, case studies of individuals who had high and low barrier scores indicated that it was the low barrier person who was more apt to have a life style which reflected impulsivity and erratic behavior. They showed patterns of marital conflict and aimless drifting as opposed to the more integrated, socially acceptable adjustment of the high barrier subjects.

In a later study, Brodie (1959) noted that subjects with low barrier scores were more inclined to be "impulsive," "uninhibited," "assertive," and "angry" while those with high barrier scores were measured as being "controlled" and "guarded."

The purpose of the present study is to further explore the relationship between the barrier score and the expression of aggression and gratification of impulses. This will be done by studying a group of juvenile delinquents confined for thefts, incorrigibility, and assaults ranging from battery to homicide. These boys, whose whole life styles represent the outward rather than inward expression of anger, as well as the free gratification of impulses, should confirm whether or not the low barrier subject is indeed more aggressive and impulsive, as it would seem from the case history data reported by Fisher and Cleveland as well as from the observations made by Brodie, or if it is, instead, the high barrier subject who is more aggressive, as it would seem from the study of the Harvard undergraduates.

SUBJECTS AND PROCEDURES

The data for this study were collected as part of a larger study on undercontrol and overcontrol among assaultive and homicidal adolescents (Megargee, 1964). As part of this study 75 male juvenile delinquents detained at the Alameda County (California) Juvenile Hall were examined by Probation Department psychologists who included in their battery of tests the Holtzman Inkblot Technique. The testing program was carried out over a 10-month period and the subjects were not informed of the fact that the results were to be used for research purposes as well as for the routine precourt psychological examination. The psychological examination differed from a routine examination only in that a standardized interview and test battery were used and all responses to projective tests, the Holtzman included, were tape-recorded and later transcribed by a stenographer.

All scoring was done by the author; later the

---

[1] The author is grateful to Carol Hampton and H. Dow Dunn who assisted in the study. This research was supported in part by the Hogg Foundation for Mental Health Research.

The data for this study were collected while the author was at the University of California and the Alameda County Probation Department Guidance Clinic.

TABLE 1

BARRIER SCORES OF 75 JUVENILE DELINQUENTS CONTRASTED WITH THOSE REPORTED FOR NONDELINQUENT SEVENTH AND ELEVENTH GRADERS (HOLTZMAN ET AL., 1961)

Group	*M*	*SD*	*N*	*t* of difference from delinquent group
Juvenile delinquents	4.2	3.3	75	—
Holtzman et al.'s (1961) 7th-grade sample	6.1	3.7	197	4.10*
Holtzman et al.'s (1961) 11th-grade sample	5.9	2.8	72	3.44*

* $p < .001$.

protocols were rescored by Carol Hampton, criterion scorer for the Holtzman Inkblot Technique research group at the University of Texas. The interscorer reliability for the scoring of the barrier variable was .86.

The 75 juvenile delinquents who were subjects in the present study ranged in age from 11–1 to 17–9 with a mean age of 15–5; their IQs as estimated from subscales of the WISC or the WAIS (depending on the subject's age), ranged from 64 to 140 with a mean of 96 and a standard deviation of 16.9; 45 of the boys were Negro and 30 were white. Thirty of the boys were detained for assaultive crimes ranging in severity from battery to homicide; 20 of the boys were detained for "incorribility," a blanket charge covering for the most part such things as unruliness and defiance in the home or school but also including drinking and, in one case, attempted intercourse with a stepmother; 25 of the boys were detained for property offenses such as automobile theft and burglary.

All the boys were examined during their first 10 days of detention while awaiting a juvenile court hearing for their offense. During this time they lived in units with 35 to 45 other boys and were supervised at all times by at least two counselors. The counselors, who ate with the boys, supervised them during sports and recreation, and were with them on a 24 hour a day basis, had an excellent opportunity to observe the interpersonal interactions and behavior of each of the boys selected for study. At the end of the third and tenth days of detention, each counselor having contact with a boy submitted a series of ratings including a five-point rating scale of aggressiveness on which a score of 1 indicated that a person, "threatens others, reacts to reproof violently, is overtly aggressive, dominant, starts trouble," while a score of 5 indicated that the subject, "complies with all requests, submits to violence without doing anything about it." They also filled out a behavior checklist which itemized 12 different types of verbally and physically aggressive behavior. Each counselor checked off all the types of aggressive behavior he had observed each boy engaging in.[2] These checklists were then used to compute scores on overall verbal and physical aggressiveness.[3]

The data were analyzed in three ways. First, mean barrier scores were computed for the delinquent group as a whole and compared with normative scores reported for nondelinquent adolescents by Holtzman, Thorpe, Swartz, and Herron (1961). Next, the delinquent sample was divided into those whose behavior was extremely delinquent and those whose delinquency was less serious. The mean barrier scores of these two subgroups were then compared. Finally, the barrier scores of the delinquent sample as a whole were correlated with the counselors' ratings of aggressiveness and the scores on overall verbal and physical aggression derived from the counselors' checklists.

RESULTS

For the first analysis, barrier scores for the delinquent sample were compared with the norms for 197 seventh-graders and 72 eleventh-graders reported by Holtzman et al. (1961, p. 125). The data were analyzed by means of *t* tests and the results showed that the delinquent group had a mean barrier score significantly lower than either of the two groups of nondelinquent adolescents (see Table 1).

While the differences between the barrier scores of delinquent and nondelinquent adolescents were consistent with what would be expected on the basis of Brodie's (1959) observations and Fisher and Cleveland's (1958) biographical data, it is possible that the lower score for the delinquents was an artifact of some other difference between the groups. One such variable is the number of words used in a response. Fisher and Cleveland (p. 68) have reported that Appleby obtained a phi coefficient of .46 between barrier scores and the number of words in total Rorschach protocols, equated for response total. While Fisher and Cleveland felt that this result was "peculiarly a function of the sample he studied . . . [p. 68]," it is consistent with the present investigator's finding of a correlation of .56 between barrier scores and the mean number of words per response in the present delinquent sample (Megargee, 1965).

Such a relationship is understandable from an examination of barrier responses. Holtzman et al. (1961, p. 75) state that barrier is

---

[2] These checklists and rating scales were devised by H. Kelley Naylor.

[3] Copies of the rating scales and checklists as well as detailed descriptions of the scoring procedure can be found in Megargee (1964).

TABLE 2

MEAN NUMBER OF WORDS PER RESPONSE AND BARRIER SCORES FOR EXTREMELY DELINQUENT AND LESS DELINQUENT JUVENILES

Group	Number of words per response				Barrier scores		
	N	*M*	*SD*	*F* of word count differences	*M*	*SD*	*F* of barrier score differences
Extremely delinquent juveniles	28[a]	26.55	12.89	1.53	3.14	3.12	4.72*
Less delinquent juveniles	44	30.57	16.42		4.75	3.02	

* $p < .001$.
[a] Data as to disposition not available for three subjects.

scored for any protective covering or surface that might symbolically relate to body-image boundaries. Frequently this means that a longer response in which such qualities are emphasized is necessary before barrier can be scored. For example, "Man with a hat" is not scored for barrier, but "man with a cap with a tassel on it" is; "skin" is not scored but "skin with stripes on it" is; "donkey" is not scored but "donkey with a load covering its back" is (pp. 75ff).

If the number of words in a response can significantly influence barrier scores, then the significantly lower barrier scores in the delinquent group, as compared with nondelinquent adolescents, could be an artifact of guardedness or decreased loquacity. For this reason, the scores of extremely delinquent and less seriously delinquent subsamples were compared after first determining that these subgroups did not differ in verbosity.

An extremely delinquent juvenile was defined as one for whom placement at the California State Youth Authority had been recommended by the investigating Probation Officer. Such placement was recommended only for those boys considered dangerous to the community or those for whom all other measures had been tried without success. Twenty-eight of the subjects fell into this category. For 44 other members of the sample less stringent measures had been recommended such as probation or placement at local county-run institutions. These boys were accordingly defined as a less seriously delinquent subgroup. (No data as to recommended disposition were available for three subjects who were accordingly deleted from this analysis.)

The fact that the inkblot tests had been tape-recorded and transcribed made it possible to count the number of words used by each subject and compute the mean number of words per response. Analysis of variance of the subgroup differences showed that there was no significant difference in verbosity between the two subgroups (see Table 2).

When the barrier scores were compared, however, it was determined that the extremely delinquent subgroup had a mean barrier score significantly lower than that of the less seriously delinquent subgroup (see Table 2).

In a final analysis, the barrier scores for the total delinquent group were correlated with scores for verbal aggression and physical aggression derived from the unit counselors' checklists, and the rating scale of aggressiveness completed by the counselors.[4] The results appear in Table 3. The barrier score was negatively but insignificantly correlated with the scores for verbal and physical aggressiveness derived from the behavior checklist. However, there was a significant positive correlation with the global rating of aggressiveness. (It

TABLE 3

CORRELATION OF BARRIER SCORES OF 74 JUVENILE DELINQUENTS WITH VARIOUS INDICES OF AGGRESSIVE BEHAVIOR WHILE IN DETENTION

Index	*r*	*p*
Verbal aggression (Behavior Checklist)	−.09	*ns*
Physical aggressiveness (Behavior Checklist)	−.17	*ns*
Rating scale of aggressiveness	+.23	<.05

Note.—Reports of behavior in detention not available for one subject because of his premature release from custody.

---

[4] One subject was deleted from this analysis since his premature release precluded obtaining these data.

will be recalled that on the rating scale a score of 1 indicated maximum aggressiveness and a score of 5 minimum aggressiveness; hence the positive correlation.) While all the correlations were of low order, they were nevertheless in the direction that would be expected on the basis of the observation that people with low barrier scores tend to be more aggressive, assertive, and impulsive than people with high barrier scores.

DISCUSSION

The data, taken as a whole, were consistent with Fisher and Cleveland's (1958) biographical studies and with Brodie's (1959) observations. The juvenile delinquent sample had significantly lower barrier scores than did the nondelinquent adolescent samples; moreover, the barrier scores of the more seriously delinquent juveniles were significantly lower than those of the less seriously delinquent boys. The latter difference was obtained despite the fact that there was no difference between the two subgroups in the mean number of words used per response. This indicated that the differences noted were not merely an artifact of guardedness or degree of verbosity in the various samples.

Correlations between barrier scores and various indices of aggressiveness of behavior during detention, while not high, were in the direction that would be expected if lower barrier scores were associated with greater aggressiveness and impulsivity. (The fact that the measures of aggressiveness were obtained in a custodial setting while the boys were awaiting a court hearing no doubt decreased the range of observed aggressive behavior.)

These results would seem to contradict those obtained by Fisher and Cleveland (1958) on Harvard undergraduates in which they found that those subjects with high barrier scores expressed anger outwardly while those with low barrier scores turned their anger inward. If, however, the barrier score is regarded as an index of adjustment or fitness rather than as an index of anger, or aggression, the contradiction becomes understandable. It is appropriate for a Harvard undergraduate who has been subjected to a deliberately provocative laboratory procedure to express anger outwardly in an interview where self-expression is encouraged. On the other hand, it is not adaptive for a juvenile delinquent to express all of his frustrations through socially disapproved aggression up to and including homicide. In both situations, then, high barrier scores were associated with adaptive behavior while low barrier scores were associated with maladjusted behavior.

There is still controversy as to whether the barrier score is indeed an index of body boundary perception (Fisher, 1963). While the present data can be interpreted without reference to the body-image concept, nevertheless, they are consistent with a body boundary interpretation of the barrier score. Erikson (1950) has pointed out that under the pressures of the physiological changes taking place at puberty, the problem of ego identity as opposed to ego diffusion assumes critical importance for the adolescent in our culture. The adolescent who has difficulty in establishing an adequate ego identity is apt to experiment with various roles including that of the delinquent. The greater his sense of ego diffusion the more he is apt to cling to such experimental roles (Witmer, 1960).

Since the problem of ego diffusion is bound up with and intimately associated with problems of changing body image, it is likely that ego diffusion would be marked by a lowered barrier score if indeed the barrier score reflects the definiteness of one's body boundary perceptions. This would account for the association between the barrier score and the degree of delinquency noted in the present study, since it would follow that those with the greatest ego diffusion (and hence the lowest barrier scores) would be likely to be the most seriously delinquent. This interpretation is also consistent with the data reported on the Harvard students (Fisher & Cleveland, 1958) since it is likely that those students with the best developed sense of ego identity (hence the highest barrier scores) would feel sufficiently independent to express their anger openly after the frustrating experimental procedures.

The patterns of normative scores reported by Holtzman et al. (1961) lend further support to this interpretation of the barrier score. They found normal adolescents by and large had mean barrier scores lower than those obtained by other nonclinical groups.[5] Among clinical samples, schizophrenics, who have a notoriously poor sense of either body boundary definiteness or ego identity, had low barrier scores. Depressed patients, whose problems are not primarily related to ego identity, had scores not only higher than those of the schizophrenics but also higher than those of the normal adolescents (Holtzman et al. 1961). These data are consistent with the notion that the ego diffusion and body image confusion of adolescence is reflected in low barrier scores and that the barrier score may be an index of ego identity as opposed to ego diffusion.

---

[5] Additional normative data collected by Holtzman and Thorpe show that a sample of 142 seventh graders, all of whom were tested within 2 weeks of age 12-8, had a mean barrier score of 5.75. Another sample of 83 sixth graders, 53 seventh graders, 43 eighth graders, and 11 ninth graders had a mean barrier score of 5.75; the mean IQ of this latter sample was 125 (Thorpe, J., personal communication, 1964). These data indicate that the low barrier scores observed in the published norms for early adolescents are quite stable and consistent.

REFERENCES

BRODIE, C. M. *The prediction of qualitative characteristics of behavior in stress situations using test-assessed personality constructs.* (Doctoral Dissertation, University of Illinois) Ann Arbor, Mich.: University Microfilms, 1959, No. 60-1619.

ERIKSON, E. H. *Childhood and society.* New York: Norton, 1950.

FISHER, S. A further appraisal of the body boundary concept. *Journal of Consulting Psychology,* 1963, 27, 62-74.

FISHER, S., & CLEVELAND, S. *Body image and personality.* New York: Nostrand, 1958.

HOLTZMAN, W. H., THORPE, J. S., SWARTZ, J. D., & HERRON, E. W. *Inkblot perception and personality.* Austin: Univer. Texas Press, 1961.

MEGARGEE, E. I. *Undercontrol and overcontrol in assaultive and homicidal adolescents.* (Doctoral Dissertation, University of California, Berkeley) Ann Arbor, Mich.: University Microfilms, 1964, No. 64-9923.

MEGARGEE, E. I. The relation of response length to Holtzman Inkblot Technique scores. *Journal of Consulting Psychology,* 1965, in press.

WITMER, HELEN L. Delinquency and the adolescent crisis. *Children's Bureau Publication No. 11.* Washington: U.S. Government Printing Office, 1960.

(Received August 17, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 4, 312–316

COMMONALITY AND STABILITY OF WORD ASSOCIATION RESPONSES IN GOOD AND POOR PREMORBID SCHIZOPHRENICS[1]

PAUL R. DOKECKI, LEWIS G. POLIDORO[2]
George Peabody College

AND RUE L. CROMWELL
Vanderbilt University Medical School

The validity of the commonality of response construct and stability of associations over time were investigated in 18 good premorbid schizophrenics (GPS), 18 poor premorbids (PPS), and 18 tuberculosis control Ss (TB). The Kent-Rosanoff Word Association Test was administered on 2 occasions separated by 48 hrs. Results showed TB = GPS > PPS on both commonality of association and stability of responses. This finding suggests the inadequacy of the notion of a universal deficiency in schizophrenic associational processes. Differences were also found between the cultural response style (C) and the idiosyncratic response style (I) on stability, with TBI = GPSI > PPSI and TBC = GPSC > PPSC. Regardless of presence or absence of psychopathology, C was greater than I on the stability measure.

The word association technique has had a long history of use in clinical and experimental studies of personality and psychopathology. Jenkins (1960) extended, both theoretically and empirically, the often observed phenomenon that two response styles can be easily differentiated on the basis of associations. Certain individuals associate to the stimulus words responses given by the majority of people. Their associations have high commonality with the culture. Other individuals tend to produce responses rarely given by most people, showing a deviation from the cultural standard. Each type is said to possess this presence or absence of correspondence to the culture in many habit systems. For the purposes of the present study these two groups were designated cultural (C) and idiosyncratic (I) responders, respectively, along a commonality continuum.

Jenkins (1960) demonstrated that the C individual is significantly more stable than the I in production of associations. The present study investigated the relevance of the commonality construct and the notion of stability as they relate to schizophrenia.

As early as Bleuler's (1950) classic work, impoverishment or deviation of the associational process has been considered a universal characteristic of schizophrenia; that is to say, schizophrenics regardless of subclassification have been said to show an across the board faulty associational mechanism. Supporting this notion Somner, Dewar, and Osmond (1960) found that schizophrenics given word association tests had less common associations and were less stable on a retest (administered 1 week after initial testing) than a nonschizophrenic control group. Lewinsohn and Elwood (1961) found support for the hypothesis that schizophrenics fail to share cultural associational expectancies. However, Shakow (1963) reported an unpublished study by Heubner conducted with the Kent-Rosanoff Test, which showed that while schizophrenics responded with fewer common associates, there was no difference in the stability of the commonality measure over time intervals ranging from 2 weeks to several months as compared with a normal control group. With a measure similar to the one used in the present study, there was a tendency for the schizophrenics to be less stable, although no statistical analysis was performed. Thus, predominant research support exists for the notion that schizophrenics universally show an associational deficit.

---

[1] Gratitude is due to William Hubbard, director of the Middle Tennessee Tuberculosis Hospital, and John Wittrig, psychologist at the Veterans Administration, Murfreesboro, Tennessee, for their cooperation in making this study possible.

[2] The present paper was written while the senior author was a USPHS Fellow in Clinical Psychology and L. Polidoro was a United Cerebral Palsy Research Fellow, both at George Peabody College, Nashville, Tennessee.

Humphries (1961) investigating the relationship between premorbid adjustment in schizophrenics and variability of associations over six 2-week testing sessions, found no significant correlation between the variables.

Studies relating the commonality construct to personality variables have not always been successful (Sarason, 1959; Terman & Miles, 1936). From a methodological point of view, Jenkins (1960) has warned that since deviation from the average on the part of the I responder could be in either the subnormal or supernormal direction, only unipolar traits classifying the average subject at the upper end of the distribution would be expected to be related to commonality; bipolar traits would not. An additional variable seems to be instructions in administering the association test. Horton, Marlowe, and Crowne (1963) found commonality under relaxed set instructions (differing from the usual speed set) to be positively related to the need for social approval. The authors suggested that previous failures to find relationships between commonality and personality may be due to the use of speed set instructions which tend to produce less revealing or diagnostically useful material. The present study took both these considerations into account.

In addition to commonality and stability of associations as they relate to schizophrenia, the present study was concerned with a subject variable, premorbid adjustment. While the Humphries (1961) study found negative results, differences have been shown to exist between the good and poor premorbid as rated by the Phillips Scale of Premorbid Adjustment (Phillips, 1953). Among the numerous variables showing reliable differences are size estimation (Harris, 1957), reminiscence (Bleke, 1955), and prognostic indications (Farina, Garmezy, & Barry, 1963; Farina & Webb, 1956). The general findings also indicated that nonschizophrenic control groups were different from both goods and poors where such comparisons were relevant.

If the schizophrenic is seen as operating in his own private world, the prediction follows that the schizophrenic associational process is less cultural than the normal, and less stable in view of the Jenkins (1960) finding. Adding the premorbid history dimension one would expect by the nature of the scale (Phillips, 1953), measuring sexual and interpersonal relations, that the good premorbid is in tune with his social context to a greater degree than the poor, and consequently has both more stable and more cultural associations. The present study investigated the hypothesis that commonality and stability of associations occur in the following hierarchy: normals > good premorbid schizophrenics > poor premorbid schizophrenics.

The notion of two distinct response styles, the I and C, was also investigated. If, as Jenkins (1960) has maintained, there are behavioral differences between the two styles in normal subjects, can comparable response styles be isolated for the good and poor premorbid groups? Further, do they behave in relatively the same manner as the two groups of normal responders?

Method

Subjects

The schizophrenic sample was composed of 36 male patients from the Veterans Administration Hospital, Murfreesboro, Tennessee. Only white, male, literate patients below age 60 who had hospital privileges and were receiving chemotherapy were included as subjects. Although the study was not specifically designed to stabilize medication over the

TABLE 1

Descriptive Data on Schizophrenic Samples

Characteristic	Group: Good premorbid $N = 18$	Group: Poor premorbid $N = 18$
Age		
$\bar{X}$	41.88	40.22
SD	3.50	6.58
Years since first hospitalization		
$\bar{X}$	8.50	12.61
SD	5.50	5.82
Years of current hospitalization		
$\bar{X}$	3.78	6.56
SD	3.32	6.01
Number of paranoids	4	7
Number of nonparanoids	14	11

TABLE 2

Means and Standard Deviations of Good Premorbid (GPS) and Poor Premorbid Schizophrenics (PPS) and Tuberculosis Control (TB) Groups on the Commonality Score (CS)

Group	$\bar{X}$	*SD*
TB	16.08	10.58
GPS	15.30	11.04
PPS	9.92	8.98

testing interval, subsequent inquiry revealed no significant changes in drug schedules for the subjects.

The primary selection variable was the score obtained on Section I of the Phillips scale (Phillips, 1953) which rates premorbid adjustment. Patients obtaining a score of 16 or above were designated poor premorbid (PPS); those obtaining 15 or below were designated good premorbid (GPS). There were 18 subjects in each of these two groups. Descriptive data on the schizophrenic samples are shown in Table 1.

Control subjects (TB) were 18 white, male, literate, tuberculosis patients below age 60 from the Middle Tennessee Tuberculosis Hospital, Nashville, Tennessee. These subjects had a mean age of 43.22 years ($SD = 14.79$). The hospital is a short-term treatment center for Tuberculosis. Although approximately 50% of the subjects were alcoholic, the group was considered a more apt control group than the typically used groups (nonpsychotic psychiatric patients, hospital personnel, or college students), due to the nonpsychiatric setting and the relative uncertainty concerning length of hospitalization on the part of the patients.

Procedure

All subjects were seen in small groups on 2 separate days with an interval of 48 hours between Day 1 and Day 2. On Day 1 subjects were given the Kent-Rosanoff Word Association Test (K-R) under relaxed set instructions (Horton et al., 1963) which did not emphasize either speed of working or the "give the first reponse which comes to your mind" aspect of the typical speed set. These instructions were used to maximize the probability of obtaining psychologically meaningful associations. Procedure was identical for Day 2.

TABLE 3

Means and Standard Deviations of TB, GPS, and PPS Groups on Stability Score (SS)

Group	$\bar{X}$	*SD*
TB	52.5	21.57
GPS	52.5	25.82
PPS	38.56	21.66

Scoring of K-R Protocols

Each K-R protocol was scored for commonality of response by a 33-item scale developed by Horton et al. (1963). These authors found the measure to correlate .97 with an overall frequency count of the 100 word list tabulated from norms (computed for relaxed set instructions). The commonality scores for Day 1 and Day 2 were combined for each subject yielding an overall mean commonality score (CS). Possible scores on CS range from 0–33 with high scores indicating high commonality.

Stability scores (SS) were computed for each subject by a tabulation of the number of associates which remained the same on the K-R protocols from Day 1 to Day 2. While change in number of a noun (singular to plural) was counted as identical responses, different verb forms were not. The range of possible scores on SS is from 0–100 with a high score indicating more stability.

TABLE 4

Means and Standard Deviations of Stability Scores (SS) as a Function of TB, GPS, and PPS Groups and Idiosyncratic (I) and Cultural (C) Response Styles

	TB Group		GPS Group		PPS Group	
	I	C	I	C	I	C
$\bar{X}$	44.56	60.44	44.00	61.00	35.22	41.89
SD	16.11	9.41	21.41	8.01	12.93	16.71

Results

Commonality Score (CS)

With CS as the dependent measure, a simple randomized design analysis of variance (Lindquist, 1953) was performed for the TB, GPS, and PPS groups. Means and standard deviations are presented in Table 2. Results were significant ($F = 3.66$, $df = 2/51$, $p < .025$). Subanalyses showed TB = GPS, TB > PPS ($t = 2.48$, $df = 51$, $p < .01$, one-tailed), and GPS > PPS ($t = 2.77$, $df = 51$, $p < .005$, one-tailed). Thus, the tuberculosis patients were not different from the good premorbids while both groups produced significantly more common associates than the poors.

While one might raise the point that the lack of TB-GPS differences might be due to the quasi-psychiatric nature of the TB group (approximately 50% alcoholics), the Horton et al. (1963) results bear mentioning. Using the identical test, instructions, and method of scoring for commonality, these authors found the mean for 153 introductory psychology students at the University of Kentucky to be 14.58. This is to be compared with the means of the present study: TB = 16.08, GPS = 15.30. As may be seen, the TB and GPS groups tended to be even more common than the college group.

Stability Scores (SS)

A first analysis was performed on broad groups (TB, GPS, PPS) with SS as the dependent measure. Means and standard deviations are presented in Table 3. The simple randomized design analysis of variance (Lindquist, 1953) was significant ($F = 3.54$, $df = 2/51$, $p < .025$). A breakdown of effects showed TB = GPS, TB > PPS ($t = 2.30$, $df = 51$, $p < .025$, one-tailed), and GPS > PPS ($t = 2.30$, $df = 51$, $p < .025$, one-tailed). Again, while there were no differences between the control and the good premorbid groups in stability over time, both were significantly different from the poor premorbids.

On the basis of the CS scores each group of subjects was split at its median yielding three idiosyncratic (I) and three cultural (C) groups: TBI, GPSI, PPSI, TBC, GPSC, and PPSC. An A × B factorial design analysis of variance (Lindquist, 1953) was performed (groups by I-C). Means and standard deviations are presented in Table 4. The A × B interaction effect was found to be significant ($F = 3.75$, $df = 2/48$, $p < .05$), requiring subanalyses of simple effects rather than description of main effects. The simple effects breakdown of each variable at every level of the other yielded a significant groups by I effect ($F = 6.39$, $df = 2/48$, $p < .005$), with TBI = GPSI, TBI > PPSI ($t = 3.19$, $df = 48$, $p < .01$, two-tailed), and GPSI > PPSI ($t = 2.39$, $df = 48$, $p < .01$, two-tailed). For the idiosyncratic groups there were no differences between the controls and good premorbids; whereas, both exceeded the poor premorbids on stability. Groups by C yielded an F of 27.58 ($df = 2/48$, $p < .001$), with TBC = GPSC, TBC > PPSC ($t = 6.33$, $df = 48$, $p < .001$, two-tailed), and GPSC > PPSC ($t = 6.52$, $df = 48$, $p < .001$, two-tailed). For the cultural groups again there were no differences between the control and good premorbid subjects; whereas, both were significantly more stable than the poor premorbids. Further simple effects showed TBC > TBI ($t = 5.42$, $df = 48$, $p < .001$, two-tailed), GPSC > GPSI ($t = 5.80$, $df = 48$, $p < .001$, two-tailed), and PPSC > PPSI ($t = 2.27$, $df = 48$, $p < .05$, two-tailed). Even across groups of pathology the Jenkins (1960) finding of cultural responders being more consistent than idiosyncratic responders was substantiated.

DISCUSSION

In agreement with the hypotheses, good premorbid schizophrenics were found to have both more common associates and a greater degree of stability over time than poor premorbids. However, contrary to expectations, both schizophrenic groups were not inferior on these measures to the control group. Only when the schizophrenics had poor premorbid adjustment were they inferior to the normals (regardless of idiosyncratic or cultural response style). Therefore, the common notion of a universal deficiency in associational processes in all schizophrenics is open to question, at least as far as correspondence to a cultural standard and stability over time are concerned. These results may help to explain the findings of Farina and Webb (1956) and Farina, Garmezy, and Barry (1963) that the schizophrenic with a good premorbid adjustment has a better prognosis for recovery than the poor premorbid. Return to the nonhospital world is probably related to the extent of correspondence of the patient's language system with that of the hospital staff (who make the release decisions) and that of the people in the patient's culture (with whom he must deal and who must deal with him). Since the good premorbid is more like the cultural standard in his associational processes than the poor, he is more likely to be in a better position to be received back into the culture.

In addition, the present study lends added validity to the commonality construct which Jenkins (1960) has proposed. Regardless of presence or absence of psychopathology the idiosyncratic responder does differ reliably from his cultural counterpart with respect to stability over time. If, as Jenkins suggests the idiosyncratic responder may deviate in either a subnormal or supernormal direction, future research could be aimed at discovering phenomena where the idiosyncratic normal differs from the idiosyncratic psychopathological individual.

REFERENCES

Bleke, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *Journal of Personality,* 1955, **23**, 479–498.

Bleuler, E. *Dementia praecox or the group of schizophrenias.* (trans. by J. Zenkin). New York: International University Press, 1950.

Farina, A., Garmezy, N., & Barry, H. Relationship of marital status to incidence and prognosis of schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **67**, 624–630.

Farina, A., & Webb, W. Premorbid adjustment and subsequent discharge. *Journal of Nervous and Mental Diseases,* 1956, **124**, 612–613.

Harris, J. G. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *Journal of Personality,* 1957, **25**, 651–671.

Horton, D. L., Marlowe, D., & Crowne, D. P. The effect of instructional set and need for social approval on commonality of word association responses. *Journal of Abnormal and Social Psychology,* 1963, **66**, 67–72.

Humphries, C. C. Temporal variability and premorbid adjustment in schizophrenia. *Dissertation Abstracts,* 1961, **21**, 2003. (Abstract)

Jenkins, J. J. Commonality of association as an indicator of general patterns of verbal behavior. In T. A. Sebeok (Ed.), *Style in language.* New York: Wiley, 1960. Pp. 307–329.

Lewinsohn, P. M., & Elwood, D. C. The role of contextual constraints in the learning of language samples in schizophrenia. *Journal of Nervous and Mental Diseases,* 1961, **133**, 78–91.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Phillips, L. Case history data and prognoses in schizophrenia. *Journal of Nervous and Mental Diseases,* 1953, **117**, 515–525.

Sarason, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *Journal of Abnormal and Social Psychology,* 1959, **59**, 37–42.

Shakow, D. Psychological deficit in schizophrenia. *Behavioral Science,* 1963, **8**, 275–305.

Somner, R., Dewar, R., & Osmond, H. Is there a schizophrenic language? *Archives of General Psychiatry,* 1960, **3**, 665–673.

Terman, L., & Miles, Catherine. *Sex and personality.* New York: McGraw-Hill, 1936.

(Received August 31, 1964)

Volume 70 October 1965 Number 5

Journal of Abnormal Psychology

Editor: Howard F. Hunt, *New York State Psychiatric Institute and Columbia University*

Associate Editor: William N. Thetford, *College of Physicians and Surgeons, Columbia University*

Consulting Editors

Theodore X. Barber
Medfield Research Foundation

Harold Basowitz
New York University

Eugene I. Burdock
New York State Psychiatric Institute and Columbia University

John M. Butler
University of Chicago

Arthur C. Carr
New York State Psychiatric Institute and Columbia University

Loren J. Chapman
Southern Illinois University

John H. Cumming
New York State Department of Mental Hygiene

W. Grant Dahlstrom
University of North Carolina

Jerome D. Frank
Johns Hopkins University

Sol L. Garfield
Columbia University

Ann Magaret Garner
Lincoln, Nebraska

Israel Goldiamond
Institute for Behavioral Research

Harrison G. Gough
University of California, Berkeley

Starke R. Hathaway
University of Minnesota

Ralph Heine
University of Chicago

Nicholas Hobbs
George Peabody College

Lawrence C. Kolb
New York State Psychiatric Institute and Columbia University

Sheldon J. Korchin
University of California

William G. Lawlor
Fordham University

Ardie Lubin
U. S. Naval Hospital, San Diego

Salvatore Maddi
University of Chicago

David Rosenthal
National Institute of Mental Health

Alan K. Rosenwald
Illinois State Psychiatric Institute

Saul Rosenzweig
Washington University, St. Louis

Jacques Rutschmann
New York State Psychiatric Institute

Seymour Sarason
Yale University

Joseph Schachter
New York State Psychiatric Institute and Columbia University

Roy Schafer
Yale University

John Shlien
University of Chicago

Morris I. Stein
New York University

Charles Strother
University of Washington

Harold L. Williams
University of Oklahoma School of Medicine

C. L. Winder
Michigan State University

Joseph Zubin
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

FRANCES L. BREWER
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 317–336

PSYCHOLOGICAL DIFFERENTIATION AND FORMS OF PATHOLOGY[1]

HERMAN A. WITKIN[2]

State University of New York, Downstate Medical Center[3]

Research on cognitive styles and the broader dimensions of personality functioning of which they are a part has important implications for personality theory and research and problems of diagnosis and therapy. This is demonstrated through the global-articulated cognitive style (manifested in perception as field-dependence-independence) and the dimension of psychological differentiation of which it is the cognitive component. Different kinds of pathology occur with impairment of integration in more differentiated and less differentiated personalities. The differentiation concept and dimensions of cognitive style may help clarify nosological problems. In therapy, persons functioning at a more differentiated or less differentiated level are likely to differ in presenting symptoms, suitability for psychotherapy, nature of relation to the therapist (transference), and prospects for change.

Recent research has demonstrated that people show characteristic, self-consistent ways of functioning in their perceptual and intellectual activities. These cognitive styles, as they have come to be called, appear to be manifestations, in the cognitive sphere, of still broader dimensions of personal functioning which cut across diverse psychological areas. The fact that these broader dimensions may be "picked up" in the person's cognitive activities, in the form of cognitive styles, has an important methodological advantage. Cognitive styles may be evaluated by controlled laboratory procedures, thereby providing an experimental, objective approach to personality study and assessment.

It is of interest, too, that the dimensions of personal functioning that have been identified through the cognitive-style work represent different ways of cutting the personality "pie" from those traditionally used. New ways of looking at personality organization are thus being suggested. Even if the personality dimensions that are now being explored, as outgrowths of the cognitive-style work, prove, in time, to be congruent with our more traditional dimensions, the outcome will be to deepen and enrich our understanding of these traditional dimensions.

The approach to cognitive activity followed in the research on cognitive styles has been a functional one. A primary concern has been with the adaptive function of cognitive processes in the psychological economy of the individual. This has led to a search for connections and consistencies across psychological areas. An outcome of this research enterprise has been to demonstrate further that the conventional categories often used in describing man's psychological life are not as separate as once believed. A by-product of the cognitive-style research has thus been its contribution to a more integrated, holistic view of personality.

In these and other ways the recent extensive research on cognitive style has significant implications for personality theory, for the methodology of personality research, and for some of the practical problems encountered by the clinician in his work on diagnosis and therapy. Some of these implications may be demonstrated by considering, as illustrative, the particular cognitive style with which we have been concerned in our laboratory, (Witkin, Dyk, Faterson, Goodenough & Karp, 1962; Witkin, Lewis, Hertzman, Machover, Meissner & Wapner, 1954).

---

[1] Portions of this paper were presented at symposium, "Approaches to Integrated Concepts of Personality Theory and Assessment," at New York Academy of Sciences under the auspices of the New York Society for Projective Techniques, June 3, 1964, and at symposium, "Clinical Psychology and Experimental Psychopathology," at International Congress of Applied Psychology, Ljubljana, Yugoslavia, August 3, 1964.

[2] The work described in this paper was supported in part by a grant (M-628) from the United States Public Health Service, National Institutes of Health.

[3] Psychology Laboratory, Department of Psychiatry.

INDICATORS OF EXTENT OF DIFFERENTIATION

Articulated versus Global Dimension of Cognitive Functioning

We first identified this cognitive style, and the dimension of personal functioning of which it is a part, in perception, where we called it "field-dependence-independence." In a field-dependent mode of perceiving, perception is strongly dominated by the overall organization of the field, and parts of the field are experienced as "fused." In a field-independent mode of perceiving, parts of the field are experienced as discrete from organized background. There is now considerable evidence that a tendency toward one or the other ways of perceiving is a consistent, pervasive characteristic of an individual's perception.

From our studies of the field-dependence-independence dimension there has emerged a variety of perceptual tests for evaluating individual differences along this dimension. In all of these tests the issue is whether or not the person is able to keep an object separate from organized field in perception. In one test, the "object" is the person's own body; in another, it is a stick; in still another, it is a simple geometric design. A brief account of these tests will make clearer the nature of the perceptual dimension they assess.

One test, the body-adjustment test, is concerned with perception of the position of the body in space. The test evaluates the person's ability to perceive his body apart from the surrounding visual field, through reference to sensations of body position. The apparatus for this test consists of a small room, which can be tilted left or right, within which is a chair, which can also be tilted left or right. The subject's task is to make his body straight while the room around him is tilted. Some persons, in carrying out this task, move their bodies into alignment with the tilted room, and in that position report that they are straight, though objectively tilted as much as 35 degrees, or even more. In this kind of performance, which we call "field dependent," perception of body position is dictated by the relation between the body and surrounding world. There seems to be a fusion between body and field in experience. At the other extreme of the performance range, we find subjects who, regardless of the position of the surrounding room, are able to bring their bodies close to the true upright. Persons who perform in this fashion —we call it a "field-independent" performance—seem to have an immediate sense of the separateness of their bodies from the surrounding world.

Another of our tests of field dependence again involves perception of the upright, but the object of perception is a neutral external object, a stick, instead of the body. The apparatus for this test consists of a luminous rod and frame, the only objects visible to the subject in the completely darkened room. With the frame tilted, the subject is required to adjust the rod to the upright. Some subjects perceive the rod as straight only when it is fully aligned with the tilted frame around it. For these field-dependent persons perception of rod position is dictated by the context provided by the axes of the surrounding frame. They cannot keep rod separate from frame; in this sense their perception is global. Other subjects, at the opposite extreme, are able to adjust the rod more or less to the true upright, independently of frame position. These field-independent persons are able to perceive a part of the field as discrete from the field; in this sense, their perception is analytical.

Still another test of field dependence, the embedded-figures test, requires the subject to locate a simple figure in a complex design which is so organized as to conceal the simple figure. For some persons the simple figure almost "pops out" of the complex design. Their perception is field independent. Others are not able to find the simple figure within the 5 minutes allowed. Their perception is field dependent.

People tend to perform in a consistent fashion in these three tests. The individual who cannot separate the simple figure from the complex embedding design also cannot keep his body apart from the surrounding tilted room or the rod apart from the surrounding frame. Going beyond the particular tests I have described, this same individual is unable to keep item apart from context in a wide variety of other perceptual situations (including such classical ones as the constancies, illusions, reversible perspective) and in situations involving other sense modalities, as touch (Axelrod & Cohen, 1961). Such consistency is indicative of a stylistic tendency in perception.

The particular stylistic tendencies we have been considering are not limited to a person's perception; they manifest themselves, in congruent form, in his intellectual activities as well. Thus, persons whose perception is field dependent do less well at solving problems which require isolating essential elements from the context in which they are presented and using them in different contexts, as, for example, the tasks employed by Duncker in his studies of functional fixity.

It is because these stylistic tendencies extend across both perception, where we are dealing with an immediately present stimulus configuration, and intellectual functioning, where we are dealing with symbolic representations, that we refer to them as *cognitive* styles. The particular cognitive style we have been considering may be described as follows: At one extreme there is a consistent tendency for experience to be global and diffuse; the organization of the field as a whole dictates the manner in which its parts are experienced. At the other extreme there is a tendency for experience to be delineated and structured; parts of a field are experienced as discrete and the field as a whole organized. To these opposite poles of the cognitive style we may apply the labels "global" and "articulated." As with the dimension of perceptual field dependence, there is no implication here that the world is peopled by two kinds of human beings. Scores for any large group on tests of this cognitive style show a continuous distribution.

I may add that a more global or more articulated quality is a stable characteristic of an individual's cognitive functioning over time. There are also consistent sex differences in the articulation-global cognitive dimension. Boys and men show greater articulation than girls and women. Small but consistent sex differences in tests of field dependence have been found with groups in the United States, in a number of western European countries (see, e.g., Andrieux, 1955; Bennet, 1956; Chateau, 1959; Franks, 1956; Wit, 1955), and in Hong Kong,[4] Israel,[5] and Sierra Leone, Africa (Dawson, 1963) as well.

Evidence that the global-articulated style of cognitive functioning is part of a still broader dimension of personal functioning has come from studies of the relation of this cognitive style to nature of the body concept, of the self and of controls and defenses.

Articulation of Body Concept

Let me consider first our studies of the body concept—that is to say, the systematic impression an individual has of his body, cognitive and affective, conscious and unconscious. Our concern has been with the cognitive, rather than the libidinal aspects of the body concept, and with the articulated-global dimension in particular. In turning to the body concept, we are in effect shifting the spotlight from experience which has its primary source "out there," our main concern to this point, to experience which has its primary source "within." There is now considerable evidence that children and adults who show an articulated cognitive style in their performance in perceptual and intellectual tasks of the kind we have been considering are also likely to have an articulated body concept—that is to say, they experience their bodies as having definite limits or boundaries and the parts within as discrete yet interrelated and formed into a definite structure.

Performance in the body-adjustment test, itself, permits some inference about articulation of the body concept. Take, for example, the person who, in order to perceive his body as upright, aligns it with the titled room, and in that position reports no experience of tilt. Such a fusion of body and field in experience, or inability to keep body and field separate, suggests a lack of clear body boundaries.

---

[4] Personal communication from Robert Goodnow.

[5] Personal communication from Martin Rothman and Joel H. Kaplan.

In another, more familiar approach to study of the body concept, we have used the figure-drawing technique. To evaluate extent of articulation of drawings of human figures, made in response to the request to draw a person and then to draw a person of the opposite sex, a 5-point sophistication-of-body-concept scale was devised. This scale does not follow the usual projective uses of the figure-drawing technique but rather considers directly characteristics of the figures drawn. Three areas of the drawings are considered in making ratings: form level, identity or role and sex differentiation, and level of detailing. In a number of studies, scores for the figure-drawing test have been shown to relate significantly to measures of cognitive style. In the drawings of field-dependent children, we find very little detail and unrealistic representation of proportioning and of body parts. Sexual characteristics are shown minimally or not at all, so that in some pairs of drawings it is difficult to tell which is male and which is female. In most cases, there is no attempt at role representation. On the other hand, in the drawings of children whose perceptual performance is at the field-independent extreme we find the body drawn in realistic proportion. Parts of the body are presented in some detail and fairly realistically. There is clear representation of sex and sex differences. Aside from indication of sex through body characteristics, the sex of the figure is also indicated by such externals as clothing. We also find attempts at role representation, suggesting a sense of the uses to which the body may be put. These differences among children in the way in which they represent the body on paper are significantly related to how they perform in the cognitive tests in the laboratory.

It may appear from the description of the kinds of drawings they make that field-dependent children, who tend to make relatively unarticulated drawings, are just not as bright as field-independent children, who tend to make highly articulated drawings. Significant correlations are in fact found between figure-drawing articulation scores and total IQ. Several studies have shown, however, that this relation is carried mainly by particular subtests of standard intelligence tests which, like the cognitive tests we use, have the task requirement of separating item from context. These subtests are block design, picture completion, and object assembly which in past factor analytic studies of the Wechsler were shown to load what we would designate an "analytical factor" (Cohen, 1957, 1959). There are only low, nonsignificant relations between figure-drawing articulation scores and scores for the Wechsler vocabulary, information, and comprehension subtests which have been shown to define a "verbal-comprehension factor." I should mention that scores for tests of field dependence, similarly, relate very highly to scores for the triumvirate of block design, picture completion, and object assembly; and they do not relate to scores for the triumvirate of vocabulary, information, and comprehension (Goodenough & Karp, 1961; Karp, 1963).

The sophistication-of-body-concept scale is easily learned and applied, even by persons without experience in the figure-drawing technique. Checks on interjudge agreement in a number of studies have shown good reliability. And, as we have just seen, scores based on the scale relate well to scores of tests of the articulated-global cognitive dimension. These characteristics of the scale, together with the ease of obtaining figure drawings from subjects, make the scale a useful assessment technique.

Let me interpolate that other studies, using more experimental means to evaluate articulation, have confirmed the relation between articulation of body concept and cognitive style (Epstein, 1957; Silverman, Cohen, Shmavonian, & Greenberg, 1961).

Sense of Separate Identity

To continue, persons with a more articulated or more global mode of cognitive functioning also differ in an important aspect of the self, namely, sense of separate identity. Persons with an articulated cognitive style give evidence of a developed sense of separate identity—that is to say, they have an awareness of needs, feelings, attributes which they recognize as their own and which they identify as distinct from those of others. Sense of separate identity implies experience of the self as segregated. It also implies experience of the self as structured; internal frames of reference have been formed and are available as guides for definition of the self. The less developed sense of separate identity of persons with a global cognitive style manifests itself in reliance on external sources for definition of their attitudes, judgments, sentiments, and of their views of themselves.

The nature of this relation may be made clearer by considering a few studies from among the many that have been done in this area.

Konstadt and Forman (1965) observed that children with a global cognitive style, when taking a test under stress and so concerned about their performance, looked up at the *face* of the adult examiner about twice as often as children with an articulated cognitive style. Similarly, Crutchfield, Woodworth, and Albrecht (1958) found that persons with a global cognitive style were relatively better at recognizing and recalling faces of people they had been with earlier. Messick and Damarin (1964) observed that field-dependent subjects showed greater incidental learning than field-independent subjects when the incidental material consisted of human faces; the relation is in the opposite direction with nonhuman incidental material (Witkin et al., 1962). These studies suggest that persons with a global style are particularly attentive to faces, the major source of cues as to what others are feeling and thinking. The reliance of persons with a global cognitive style on external sources for self-definition was demonstrated in quite a different way by Linton (1955). She found that in an autokinetic situation, such persons more often changed their judgment about movement of the point of light in conformance with the suggestion of a planted confederate. The results of these and numerous other studies may be summarized by saying that the person for whom the frame around the rod, or the room around his body, strongly influence the manner in which rod and body are experienced, is, similarly, strongly influenced by the immediate social context in his experience of himself.

I would like to mention here a recently completed study by Winestine (1964) on the "twinning reaction" of boy twins because of the particularly direct way in which sense of separate identity was evaluated. The twinning reaction, assessed by interview, was rated as strong for a given twin if in his specific attitudes, feelings, and actions he showed that he experienced himself as an integral part of the twinship rather than as individuated. Twins rated high on the twinning reaction were strikingly more field dependent and their representation of the body in their figure drawings much more global. In fact, so specific a characteristic as whether the twins dressed alike or differently proved to be significantly discriminating with regard to both perceptual field dependence and nature of body concept in figure drawings. It would be of interest to study the parents of twins in relation to strength of the twinning reaction. We might expect that mothers of twins who show a weak twinning reaction and who are field independent to perceive their twins as separate individuals, rather than as a unit; to have a distinct relation with each child; and to foster the separation of the twins. These expectations are based on an earlier finding that mothers of more field-independent children, in their interactions with them, give their children greater opportunity for separation (Witkin et al., 1962).

Differences between field-dependent and field-independent persons, as a function of differences in sense of separate identity, have even been observed in characteristics of their dreams. In an experimental study of dreaming, using rapid eye movements and Stage-1 EEG sleep as indicators of the dream state,[6] observed that field-dependent subjects more often dreamed overtly about the laboratory situation. Moreover, in a recent analysis of the data from that study we found a difference between field-dependent and field-independent subjects in the kinds of incorporation they made. Field-dependent subjects more often had dreams concerned with their relation to the experimenter. Apparently these subjects get "caught up" more with another person in the laboratory situation, as we have found they do in general, and these feelings are carried over into the dream.

---

[6] Unpublished manuscript from Helen B. Lewis, D. R. Goodenough, A. Shapiro, and I. Sleser, 1965.

Specialization of Defenses

Finally, let me comment on the relation between cognitive style and nature of defenses. Studies have shown that persons who experience in articulated fashion tend to use specialized defenses, as isolation. In contrast, persons with a global cognitive style tend to use such defenses as massive repression and primitive denial. These latter defenses involve an indiscriminate, total blotting out of memory for past experiences and of perception of stimuli. Compared to such mechanisms as isolation, they represent relatively nonspecific ways of functioning.

The contrasting kinds of defenses used by persons with a more global or more articulated cognitive style may be conceived in terms similar to those we used earlier in characterizing their cognitive functioning. In the last analysis, defenses help determine the content of a person's experience—what enters into consciousness and what is put aside. They do this, in part, through regulating the interrelation between affect, on the one hand, and ideation and perception on the other. It seems true of persons with a global cognitive style that feelings strongly influence thought and perception, in other words, that feelings are not kept sufficiently discrete from thoughts and percepts. This is congruent with what we saw happen within their perception, where again they are unable to "keep things separate"—as body separate from field, rod separate from frame, or simple figure separate from organized ground. Persons with an articulated cognitive style, in their use of isolation, maintain the discreteness of feelings and ideas, although the feeling component may be "split off."

This view of communality in mode of functioning in the areas of cognition and of defenses may be made more evident by considering a few representative studies.

Bertini (1961) carried out a study of the relation between cognitive style and defenses, following this view. He considered that the "capacity" to separate and isolate an idea from its emotional content, involved in the mechanism of isolation, parallels the capacity in field-independent perception to "separate several elements from the phenomenal field in isolating them from a context." To assess the use of isolation as a defense, Bertini relied on the Rorschach, basing his analysis on the work of Schafer (1954). The expectation that the tendency to use isolation would go with field-independent perception was confirmed.

A recent study by Minard[7] on perceptual defense effectively complements Bertini's study. Words matched in structure, and previously found in a free-association test to be neutral or charged for the particular subject, were presented tachistoscopically. Persons who were field dependent showed a considerable perceptual-defense effect; their speed of perception of words was markedly affected by whether or not the word carried an emotional connotation for them. Percept and feeling were, in other words, not kept separate. Field-independent persons showed no difference in speed of perception of neutral and charged words, suggesting discreteness of percept and feeling. In their use of denial and repression as characteristic defenses, persons with a global cognitive style are in effect showing a particularly extreme influence of feeling on percepts and memories. In these instances the total experience, including both its cognitive and affective components, is "split off."

A particularly striking example of this kind of complete blotting out of experience by field-dependent persons is found in their tendency to forget their dreams, presumably because of their use of repression as a characteristic defense. A connection between field dependence and dream recall has been found by Eagle[8] in one study of 10-year-olds and in another study of 17-year-olds. It has been observed by Linton[9] and Schonbar (1964) with groups of college students. More recently it was found again in an analysis we made of data from the study of Lewis et al. (see Footnote 6). Eight subjects at each extreme of the field-dependence dimension were picked from their total group of 46 college students. Considering frequency of dream reports in a home dream diary these subjects kept, seven of the eight most field-dependent subjects and only one of the eight most field-independent subjects were found to be "non-reporters"—that is, failed to recall dreams. Recent evidence from studies using rapid eye movement and Stage-1 EEG sleep as indicators of dreaming makes it entirely clear that everyone dreams a number of times each night (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957). A lack of dream reports is therefore indicative of a failure to recall dreams rather than a failure to dream.

---

[7] Personal communication from James Minard.

[8] Personal communication from Carol Eagle.

[9] Personal communication from Harriet Linton.

The hypothesis that the failure of field-dependent persons to remember their dreams is, in part at least, a function of repression, is being tested in a study Donald Goodenough, Helen Lewis, Arthur Shapiro and I are carrying out. This study is making use of a technique Helen Lewis and I developed in an earlier preliminary study (Witkin & Lewis, in press). The technique consists of creating an important psychological event for the person just before he goes to sleep and obtaining his subsequent dreams by awakening him during each period of rapid eye movement and Stage-1 EEG. The presleep event we use consists, on some occasions, of viewing an emotionally charged film; on other occasions, viewing a neutral film; and on still other occasions, an encounter with another person through the medium of suggestion. The results of the preliminary study showed, first of all, that the presleep event often found expression in the subject's subsequent dreams in a clearly identifiable way. Further, reports from subjects that they had been dreaming, but could not remember the dream, occurred more often on awakenings following the exciting presleep event than on awakenings following the neutral presleep event. We anticipated this outcome on the premise that the exciting presleep event, by "charging" subsequent dreams, would make these dreams better "candidates for repression" and so reduce dream recall. In the study Goodenough, Lewis, Shapiro, and I are now doing we are following this same procedure of presleep stimulation, using field-dependent and field-independent persons as our subjects. It is our expectation that field-dependent persons will be more likely to "lose" dreams charged by an exciting presleep experience. Such an outcome would lend further support to the view that field-dependent persons tend to use repression as a typical mode of defense.

Summary of Indicators of Differentiation

Reviewing the evidence considered to this point, a tendency toward a more global or more articulated cognitive style has been shown to be associated with differences in body concept, in sense of separate identity, and in nature of defenses. It is now our view that the characteristics which make up the contrasting constellations described may be conceived as diverse manifestations of more developed or less developed psychological differentiation. Thus, we consider it more differentiated if, in his perception of the world, the person perceives parts of the field as discrete and the field as structured. We consider it more differentiated if, in his concept of his body, the person has a definite sense of the boundaries of the body and of the interrelation among its parts. We consider it more differentiated if the person has a feeling of himself as an individual distinct from others and has internalized, developed standards to guide his view of the world and of himself. We consider it more differentiated if the defenses the person uses are specialized. It is our view that these various characteristics, which we have found to cluster together, are not the end-products of development in separate channels, but are diverse expressions of an underlying process of development toward greater psychological complexity. "Level of differentiation" is a concept which encourages us to look across psychological areas and provides a basis for thinking about self-consistency in individual psychological make-up.

The level of differentiation at which a person functions may be assessed in different areas. For clinical evaluation one or several areas may be considered. It is perhaps best assessed through the person's cognitive style, where, as noted earlier, objective, experimental means of evaluation may be used. In this sense a person's cognitive style, in the articulated-global dimension, may be considered a "tracer element." Certainly, assessment of psychological differentiation need not be limited to the cognitive sphere. It may also be assessed effectively, for example, in the area of the body concept, by means of the figure-drawing sophistication-of-body-concept scale which I described earlier. In our own work, when we want to make a rapid assessment with regard to differentiation we use the embedded-figures test and figure-drawing test, both of which are easily administered and scored, but time permitting, we like to use the rod-and-frame test as well. If the Wechsler has been given, we use the score for the block-design subtest, which, as we saw, is very similar in its requirements to the tests of field dependence and is in fact a good test of field dependence.

DIFFERENTIATION AND PROBLEMS OF PATHOLOGY

The concept of differentiation and the techniques devised to evaluate level of differentiation appear useful in both the conceptualization and the assessment of normal personality functioning. They have relevance to pathological functioning as well.

To make this clear, we need to consider the relation between differentiation and integration. Differentiation refers to the complexity of structure of a psychological system. One of the main characteristics of greater differentiation is specialization of function; another is clear separation of self from nonself. Integration refers particularly to the form of the functional relations among parts of a psychological system and between the system and its surroundings. At any level of differentiation varied modes of integration are possible, although more complex integrations may be expected with more developed differentiation. Adjustment is mainly a function of effectiveness of integration—that is, a more or less harmonious working together of parts of the system with each other and of the system as a whole with its environment.[10] Adequate adjustment is to be found at any level of differentiation, resulting from integrations effective for that level, although the nature of adjustment that may be considered adequate varies from level to level. Moreover, impaired integration, with resulting pathology, may also occur at all levels of differentiation. However, and this is of particular importance, impairment is likely to take different forms in relatively more differentiated and less differentiated personalities.

The evidence now clearly indicates that pathology occurs at both extremes of the differentiation dimension. In fact, there is some suggestion of greater frequency of pathology at the extremes than in the middle of the range. Further, pathology takes quite different forms at the two extremes. And the kinds of pathology that have been found at each extreme may be conceived as having the form which impaired integration is likely to take when more differentiated or less differentiated personalities break down. Let me illustrate this.

When personality disturbances occur among persons with a global cognitive style, and with other characteristics of limited differentiation associated with it, we are likely to find severe identity problems, with little struggle for maintenance of identity; symptoms often considered suggestive of deep-seated problems of dependence; inadequately developed controls, resulting in chaotic functioning; and passivity and helplessness. Several studies have demonstrated marked field dependence in clinical groups with symptoms commonly regarded as rooted in severe dependency problems, or in what we earlier called a lack of developed sense of separate identity. Alcoholics have been found to present a consistent picture of marked field dependence (Bailey, Hustmyer, & Kristofferson, 1961; Karp & Konstadt[11]; Karp, Poster, & Goodman, 1963; Karp, Witkin, & Goodenough, 1965a; Karp, Witkin, & Goodenough, 1965b; Witkin, Karp, & Goodenough, 1959). This picture has been found among alcoholic men and alcoholic women; among current drinkers and among abstaining former drinkers; among alcoholics who are under the influence of alcohol and among alcoholics who are sober; among long-term drinkers and among relatively new drinkers. As another indicator of their limited differentiation, alcoholics also show a relatively global body concept in their figure drawings.

---

[10] Poorly developed differentiation may sometimes be a main source of psychopathology, as, for example, when the nervous system is inadequately formed. Such impairment hampers the development of the kinds of integration needed for adjustment to many life situations.

[11] Personal communication from Stephen A. Karp.

In addition to alcoholics, other clinical groups commonly conceived to suffer severe dependency problems, which also show marked field dependence, are ulcer patients (Gordon, 1953), obese people (Pardes & Karp, in press), and asthmatic children (Fishbein, 1958).

Other kinds of clinical groups observed to show a global cognitive style have included patients with an hysterical character structure (Zukmann, 1957); patients with character disorders—that is, generally inadequate personalities unable to manage the ordinary problems of living, patients who somatize their complaints and deny any psychological problems, patients whose primary symptom is affective discharge, rather than defensive symptom organization (Korchin[12]); patients with functional cardiac disorders (Soll, 1963); and catatonics (Janucci, 1964).

Taylor (1956) has shown that psychotics who hallucinate are more likely to be field dependent, as compared to delusional psychotics, who tend to be field independent. He predicted this outcome on the ground that hallucinatory states imply dissolution of ego boundaries, whereas delusional states represent attempts to maintain separate identity and ego integrity. Compared to hallucinations, delusions have a more logical structure and they do not represent as gross a fusion between self and nonself. In hallucinations, inner states often become indistinguishable from reality, reflecting reduction of ego-integrity and of sense of individuality to an infantile stage. Taylor's finding has recently been confirmed by Powell (1964).

A recent experimental study (Bertini, Lewis, & Witkin, 1964) in which we sought to induce a hypnagogic reverie state in normal subjects yielded results consistent with Taylor's. White noise was fed into the subject's ears and halves of ping-pong balls were placed over his open eyes creating a homogeneous visual field. Under these circumstances the subject was asked, while lying down, to keep on talking continuously, describing any thoughts, feelings, or images he might have. White noise, which creates a continuous monotonous sound, has been generally known to facilitate drowsiness. It also reduces feedback from the subject's own voice and distorts the quality of the voice. A homogeneous field facilitates imagery and makes the images more plausible. Our subjects were college students, selected as being extremely field dependent or extremely field independent. Each subject was tested first under the conditions described and on another occasion immediately after viewing an exciting film. Preliminary observations suggest, as predicted, that field-dependent and field-independent subjects react quite differently to this experimental-hypnagogic procedure. Particularly relevant to the Taylor finding is the observation that field-dependent subjects more often reported vivid imagery. However bizarre the imagery, it had the quality of "actually being there," rather than of "being imagined." The reports of these field-dependent subjects suggest that the experience was one of being caught up in an ongoing scene. Of course, in contrast to Taylor's hallucinating patients, our subjects who had this imagery experience were able to tell it was not "real" even when it appeared very vivid to them. Another observation was that the hypnagogic reverie of field-dependent subjects, after viewing the exciting film, tended to be flooded with symbolic references to the film. The field-dependent subjects were also made more uneasy and even frightened by the experimental situation. Silverman, Cohen, Shmavonian and Greenberg (1961) have also reported greater "ego weakening" in field-dependent subjects under conditions of sensory isolation; and they report, as do Holt and Goldberger (1959), that field-dependent subjects more often experience imagery which appears to have an "outside" source. The indication that hallucinatorylike experiences are more readily induced in field-dependent persons, with accompanying feelings of severe discomfort, reflects on their weak ego boundaries.

Turning to the kinds of pathology encountered among differentiated persons, when they break down, we find that they tend to show delusions, expansive and euphoric ideas of grandeur, outward direction of aggression, overideation, and continuing struggle for the maintenance of identity, however bizarre the attempt. An articulated cognitive style has been found among paranoids (Janucci, 1964; Powell, 1964; Witkin et al., 1954), obsessive-compulsive characters (Zukmann, 1957), neurotics with organized symptom pictures, and those ambulatory schizophrenics who have a well-developed defensive structure (Korchin [13]).

---

[12] Personal communication from Sheldon Korchin.

It is not easy to work out precisely the linkages between extent of differentiation and forms of pathology. For one thing, the pathways by which particular kinds of pathology may emerge in a given personality setting are inordinately complex and difficult to trace; for another, level of differentiation is only one determinant, though an important one, of the course that pathological development may take. Moreover, present knowledge of the dynamics of various kinds of pathology is uncertain. Even with these difficulties, it seems possible to see, for at least some of the pathological states listed, why they may be expected in a setting of relatively developed differentiation, on the one hand, or limited differentiation, on the other.

As one illustration, we may speculate about some of the ways in which the characteristics of limitedly differentiated persons may make the development of obesity more likely among them. Their inadequately developed sense of separate identity makes it plausible that under stress they would seek comfort in oral activities that had been an important source of satisfaction in the period of close unity with mother. As a technique of defense for dealing with anxiety, eating is a nonspecialized defense. It is applied indiscriminately in a wide range of stressful situations, and it does not act in a specific, directed fashion upon the source of stress. In particular kinds of persons, it may suffuse the organism with an animal pleasure which blurs anxiety.

We may think along similar lines about alcoholism, a condition which, in its dynamic aspects, is often considered equivalent to obesity. The nonspecialized character of the defense which the use of alcohol represents is particularly clear. Alcohol is resorted to when stress becomes too great, regardless of the source of the stress, and it seems to affect experience in general, not only the part responsible for the anxiety. If the drinking is carried far enough, the self, which to begin with is limitedly differentiated, may be temporarily obliterated.

Obviously, a symptom like alcoholism is only one of many possible "choices" of pathology open to the relatively undifferentiated person when integration is impaired. Reflecting the possibility that the same symptom may be achieved by different dynamic routes, we must note that alcoholism may also sometimes, though rarely, be the choice of relatively differentiated persons. Although we have found that alcoholics who perform in a field-independent fashion are quite uncommon, their very existence points to the need for seeking other routes to the development of alcoholism than the one considered. Cognitive style may be a means of differential diagnosis of these different kinds of alcoholics. And it is possible that through the cognitive style approach other presumably homogeneous diagnostic categories may be broken down and diagnostic classification thereby refined.

A final illustration may be found in patients in whom paranoid reactions are central in the symptom picture. In this patient group, an articulated cognitive style is frequently found. Projection, a characteristic defense of the paranoid, is quite specialized, in comparison to such generalized tension-reducing techniques as eating and drinking. The paranoid projects his own system of ideas upon the world, and does so in a highly selective fashion—particular people, particular situations may be especially implicated. Such selectivity requires that experience of the world be articulated. In this connection, the paranoid is noted for his detailed, articulated system of ideas. As an attempt at preservation of the self, projection contrasts with the alcoholic's preferred way of dealing with stress, which in extreme cases results in the dissolution of the self in drink. The use of projection as a device for self-preservation, however bizarre, presupposes a self that has achieved some degree of differentiation.

These illustrations, though sketchy, suggest some of the ways in which level of differentiation may help determine form of pathology.

---

[13] Personal communication from Sheldon Korchin.

Although cognitive style, and associated characteristics of more developed and less developed differentiation, relate to particular symptoms and symptom pictures, they do not relate to major conventional nosological categories, as neuroses and schizophrenia (Bound, 1951; Franks, 1956). Among hospitalized psychiatric patients we found cases diagnosed "schizophrenic" scattered throughout the distribution of measures of perceptual field dependence (Witkin et al., 1954). Bennett (1956) and Bailey et al. (1961) have reported similar results.[14] There is no contradiction in the finding that cognitive style relates to symptoms but not to conventional nosological categories. Nosological categories are, in varying degree, based on symptoms, dynamics, and etiology. There is further a lack of consistency with which nosological categories are applied in assessment. On the other hand, to the extent that a given symptom is the end-product of particular dynamic processes, it may serve to identify these processes. Classification in terms of particular symptom pictures is therefore likely to bring together persons with common underlying dynamic processes. Although symptoms are the main basis of classification in some diagnostic categories, in many instances classification on the basis of symptoms may transcend common diagnostic categories. For example, depression may be a major symptom regardless of whether the overall diagnosis is neurosis, schizophrenia, reactive depression, etc.

The finding that cognitive style, which mirrors a deep aspect of psychological makeup, cuts across some of the conventional nosological categories is consistent with the growing shift in clinical classification from an emphasis on behavior to an emphasis on dynamic characterization. Cognitive style may thus offer one potentially useful basis for clarifying nosological problems.

That cognitive style may relate to deeper levels of functioning, and so go beyond manifest behavior, is indicated in an interesting way in a study by Gordon (1953). Gordon found ulcer patients to be markedly field dependent; and he also found significant agreement between their field-dependence scores and ratings made of them by their physicians on a specially devised dependency scale. However, ratings of dependence by the patients themselves, using the same scale, did *not* relate to their field-dependence scores. (Let me interpolate here that, for a comparison group of neurotics, field-dependence measures, physician's ratings of dependence and self-ratings of dependence were all significantly interrelated.)

In interpreting these findings for the ulcer patients, it is of special interest that on the dependency scale they rated themselves as more independent than did their physicians. The direction of the discrepancy between self-ratings of dependency, on the one hand, and physician's ratings and field-dependence measures, on the other hand, seems to suggest that, although tending to be very dependent, these patients viewed themselves as relatively independent. A common clinical view of the ulcer patient is of an overstriving person who, through his striving, is seeking to compensate for deep-seated passivity and persistent dependency needs (see for example, Alexander, 1950). He seeks to make himself appear, both to himself and to others, as more active than he is, and as capable of functioning independently of external support. Such an interpretation would be consistent with Gordon's results. It is thus possible that an individual's cognitive style may "penetrate" his apparent assertiveness and reflect the strong underlying passivity and need for external support.

Findings such as these suggest again that assessment of cognitive style may, in particular instances, be of aid in differential diagnosis.

The differentiation concept seems to provide a useful approach to the study of differences among people in directions of pathological development and in choice of symptoms. At the present time, however, it is more useful in understanding the basis of differences in general classes of pathology and symptom formation than in understanding of "choices" within a general class. For example, as we saw, both alcoholics and obese people tend to be extremely field dependent. The question remains as to why some limitedly differentiated people choose drinking as their main symptom while others choose overeating. To answer this question, intensive study of individual cases, showing such symptoms, is required, with attention given to specific patterns of integration, to particular content aspects of personality, and to nature of life circumstances. As an illustration of how these may be important, the suggestion has been made that an individual's personal values, derived from his cultural group, may place a taboo on drinking and so make overeating a more likely symptom choice.

---

[14] Cognitive style does, however, seem to relate to kind of schizophrenia. As noted, paranoids tend to be field independent (Janucci, 1964; Witkin et al., 1954). On the other hand, catatonics tend to be field dependent (Janucci, 1964). Bryant (1961) has found process schizophrenics to be significantly more field dependent than reactive schizophrenics, although this finding was not confirmed by Cancro (1962).

I have tried in this discussion of pathology to suggest that the characteristics of relatively greater or more limited differentiation may play a role in channeling pathological development. It follows from this view that the techniques devised to evaluate level of differentiation in various areas—and particularly in the area of cognitive functioning—may contribute to a more discriminating assessment of psychopathology.

The cognitive style approach has promise for the study of another kind of pathology—mental retardation. Its value is demonstrated by the results of a study we recently completed with two groups of high grade retarded boys (Witkin, Faterson, Goodenough, & Birnbaum, in press). One group consisted of "teachable" institutionalized retarded boys in the 60-80 IQ range; the other group consisted of boys in the same IQ range but living at home and attending public school classes for the mentally retarded. All the boys were given the Wechsler, tests of perceptual field dependence and the figure-drawing test. We found, first of all, that the pattern of relation among test measures in both retarded groups was similar to that previously observed in normals. Thus, measures of perceptual field dependence for the retarded boys were related to pro-rated "analytical IQs" (based, it will be recalled, on the Wechsler block design, picture completion, and object assembly subtests); they did not relate to pro-rated "verbal-comprehension IQs" (based on the vocabulary, information, and comprehension subtests). Again paralleling previous findings for normals, measures of analytical competence both from tests of field dependence and from the Wechsler scales were related to figure-drawing measures of articulation of body concept.

A second finding of interest was a marked and pervasive discrepancy in level of functioning between the analytical and verbal-comprehension clusters. Thus, on the Wechsler scales, almost all the retarded boys showed a pattern of extremely poor performance on the triumvirate of verbal-comprehension subtests and relatively much better performance on the triumvirate of analytical subtests. The discrepancy in cognitive functioning that seems to exist in these retarded boys is much sharper when the verbal-comprehension and analytical pro-rated IQs are compared than when the usual Verbal and Performance IQs are compared. Thus, in the group of institutionalized retarded boys the pro-rated verbal-comprehension and analytical IQs were 61.2 and 80.8, respectively, whereas the Verbal and Performance IQs were 71.1 and 76.3.

The frequency of occurrence of cases with relatively high analytical ability and relatively low verbal-comprehension ability in groups now identified as retarded may be the result of "routing" of children on the basis of particular emphasis upon verbal skills. If we consider children as nature made them, rather than as they are defined by Wechsler performance, we may expect to find a great many different kinds of cognitive patterns among them, with the possibility as well of deficits in one or more cognitive areas. Children with a deficit in verbal skills, whatever other cognitive strengths they may have, are especially likely to come to attention in the family, in society and in school; and on standard intelligence tests they are likely to earn a low IQ because of the emphasis on subtests involving verbal skills in the total IQ. On the other hand, we may speculate, children with other kinds of cognitive deficit than a verbal one, and who function at an overall level similar to those now commonly classified as retarded, do not come to attention as readily. Though not classified as retarded by the standards currently used in our society, such children may be as handicapped in their cognitive and personality functioning as those now labelled "retarded," but in different ways.

A cognitive-style approach to the understanding and assessment of the retarded has clear advantages over an IQ approach. In particular the cognitive-style approach offers a more comprehensive and complex view of cognitive functioning than does the IQ. It has a more developed conceptual rationale, it encompasses broader segments of cognitive functioning, and it recognizes the rooting of intellectual characteristics in personality.[15]

DIFFERENTIATION AND PROBLEMS OF THERAPY

Problems of assessment arise in a particularly complex and interesting way in the area of therapy. Persons functioning at a more differentiated or less differentiated level may be expected to differ predictably with regard to their attitudes toward psychotherapy, and their performance in the psychotherapy situation, in ways that are likely to affect the course of therapy. I would like to consider several of the specific areas in which differences may be anticipated. In some of these areas, studies are under way that I can tell you about; for other areas I can only specify the issues involved.

One important area is the nature of presenting symptoms. Our earlier discussion of forms of pathology in relation to extent of differentiation is of course relevant here.

A second area is suitability for psychological forms of therapy. Several studies have shown that therapists are not likely to accept field-dependent patients for psychotherapy; they are more apt to recommend them for somatic forms of therapy. There thus seems to be considerable agreement between the performance of a particular patient on the rod-and-frame test and the judgment independently made by the therapist about that patient. The significance of this finding needs to be explored further; its pursuit may help us in identifying more precisely the criteria used in decisions about suitability for therapy, and in determining the validity of these criteria.

A judgment that needs to be made as a prelude to deciding whether a given patient is suitable for psychotherapy is the nature of the problems he is likely to present in the course of therapy. There can be little doubt, from the characterizations given earlier, that more differentiated and less differentiated patients are likely to differ in the difficulties encountered with them. But it is hard to accept the idea that the one kind of patient is amenable to psychotherapy and the other not. Even if there is some difference between them in this respect, it may very well be that each will benefit from quite different treatment procedures. Perhaps the core issue is not suitability for psychotherapy, but which form of psychotherapy for which kind of patient. Focusing on the kinds of therapy that may help persons who show the characteristics associated with field dependence may aid us in finding ways of treating at least some of the field-dependent patients who now seem to be rejected en masse.

---

[15] Another illustration of the usefulness of the cognitive-style approach in understanding and evaluating groups with developmental deficits is found in a study we made of blind children (Witkin, 1965). Using nonvisual tests of perceptual field dependence and of analytical functioning in problem solving we were able to demonstrate that blind children, like normals, show a wide range of individual differences along the global-articulated cognitive dimension. It is of special significance for the problem of assessment that children of the same level on the verbal components of the Wechsler, to which testing of the blind is necessarily limited, differed in the extent to which their performance on our cognitive test battery showed a more global or articulated style. Assessment of the blind has, understandably, placed great emphasis on verbal kinds of performance, resulting, in some cases, in an unrepresentative and even misleading picture of their cognitive make-up. The possibility of adding tests of cognitive style to existing test procedures offers the promise of a more comprehensive cognitive evaluation of the blind. Thus, the demonstration that measures of perceptual field dependence relate highly to Wechsler analytical subtest scores, which cannot be obtained from the blind, means that through nonvisual tests of field dependence, measures of this area of Wechsler performance may be obtained. Moreover, to the extent that an individual's cognitive style carries a message about his personality functioning, opportunities for assessing personality, which are again severely limited in the blind, may be opened up.

As a related point, we have the impression—and it is only an impression at this point—that field-dependent persons are less likely to present themselves as candidates for psychotherapy. I am sure that this relation, if substantiated, will turn out to have quite a complex basis. One factor that may be involved is the tendency of field-dependent persons, noted earlier, to somatize their difficulties. On this basis, we would be more likely to pick up psychologically disturbed field-dependent patients at medical than at psychiatric clinics. Another factor may be that relatively undifferentiated persons are not as active in seeking help of an out-patient kind.

A third aspect of psychotherapy in which more differentiated and less differentiated patients may differ is in the nature of their relation to the therapist—that is, in the transference. Freedman (1962) has observed that field-dependent patients are likely to "feel better" earlier than field-independent patients, although the improvement is not maintained. We are now doing a study to check the idea that this difference is a function of differences in manner of relating to the therapist. We may speculate that the easier "fusion" of the less differentiated patient with another person, dictated by his limited sense of separate identity, manifests itself in his relation to the therapist as we know it does in his everyday relations with people. This fusion represents, in salient ways, a return to an earlier state—that of unity with mother—which, as our studies of family background suggest, such persons may not have outgrown. The reestablishment of this state may itself make such persons feel better without any real problems having been worked through. In addition to this transference effect, it may be expected that less differentiated patients are likely to be more accepting of the therapist's suggestions and interpretations, again with the consequence that they would feel better. On the other side, relatively differentiated patients are apt to be more cautious in the development of their relation to the therapist, and to filter the therapist's suggestions through their own structured systems of feelings and values.

In a study Helen Lewis, Edmund Weil, and I are now doing we are examining the development of the patient-therapist relation, in a small group of patients, in the course of the first 20 sessions of therapy. Our patients were at the extremes of the differentiation dimension, selected from a large out-patient clinic population who received our tests as part of the initial clinic screening. Each participating therapist was assigned one highly differentiated and one relatively undifferentiated patient. The therapist knew nothing of our findings on his patients. All therapy sessions were tape-recorded.

The data-gathering phase of this study has been completed, but the microscopic analysis of the records, required to check out hypotheses about differences in patient-therapist relations at the two extremes, is still under way. The kinds of specific issues we are examining in our study of the transcripts of the therapy sessions, bearing on these hypotheses, include the patient's expectations of the therapist—for example, "you will make me better" versus "you will help me figure things out so I can think straight"; the patient's reaction to the therapist's suggestions—for example, whether readily accepting or circumspect; the speed with which the patient establishes an overt bond with the therapist; the patient's ability to keep separate the transference and the reality of the doctor-patient relationship.

Though the focus in this study is on the patient's contribution to the patient-therapist relation, inevitably we are dealing with an interaction to which both participants contribute. That characteristics of the therapist himself are important and need to be considered is demonstrated in an interesting way in a study by Pollack and Kiev (1963). This study showed that psychiatrists who tested extremely field independent on our perceptual tests tended to favor either a directive and instructional or a passive observational approach to their patients in therapy, whereas relatively more (though not extremely) field-dependent therapists favored personal and mutual relations with their patients. The importance of considering the patient-therapist relation in interaction terms is indicated by an observation we made of frequency of comments by the therapist in the course of a therapy session. Therapists made strikingly more comments with limitedly differentiated patients than with highly differentiated patients (in the first session, for example, about three times as many); but at the same time therapists seemed to differ among themselves in frequency of making comments, whatever kind of patient they were treating.

Obviously, for each comment by the therapist, made at his initiative, there is a patient comment in reaction, and vice-versa. Accordingly, another way of stating the finding on frequency of comments is that there are many more patient-therapist interactions per session on the side of relatively undifferentiated patients than on the side of differentiated patients. An analysis we made of transcripts of the first therapy session indicates a variety of reasons for this difference, reflecting differences in interaction that might be expected from already known differences between more differentiated and less differentiated persons. Thus, among undifferentiated patients we find some of whom the therapist had to ask many specific questions in order to obtain the information he wanted, and the patient's responses were typically very brief. The outcome, of course, was a large number of interactions per session. Other undifferentiated patients showed a state of diffuse excitement, spurting out in a pressured way accounts of their pervasive difficulties, troubled feelings, and inadequacies.[16] Though in contrast to the first kind of undifferentiated patients these patients "pour out" a great deal, what they tell often comes in eruptive spurts, involves a great deal of repetition, and requires "containment" by the therapist. To obtain the information he wants, the therapist may still need to ask many questions, again making for many interactions.

Obviously, with both the "teeth-pulling" behavior and the "spilling" behavior observed on the undifferentiated side, considerable participation by the therapist is required for the therapist to obtain the information he wants about the patient's difficulties. In contrast, differentiated patients typically came to the first session with an articulated account of their problems and their ideas about the sources of these problems, whether dynamically correct or not. Sometimes their "opening presentation" gave the impression of a prepared statement. One of these patients actually brought with him a letter he had written to his wife about himself, which he proceeded to read to the therapist. Because such presentations may go on at great length, sometimes even bypassing the therapist's occasional attempts at intervention, the result is a small number of total interactions per session.

Another characteristic of some undifferentiated patients which contributes to a large frequency of interactions is their ready acceptance of the therapist's suggestions, sometimes in a single phrase of agreement ("You're right, doctor.") or a repetition of part of the therapist's statement. The tendency among more differentiated patients to qualify their agreement with the therapist's proposals or to disagree with them, sometimes at great length, of course makes for a smaller total number of interactions in the session. Another feature of the behavior of some undifferentiated patients contributing to a greater frequency of interactions is their attempt to prolong the hour out of feelings of separation anxiety; this has even been observed in the very first hour. An occasional differentiated patient may also resist the ending of the session, but where this has been found it was because the patient was interrupted in the midst of a developed account he was presenting and which he was intent on completing. Still another phenomenon observed in some undifferentiated patients, which again increases number of interactions, is their repeated solicitation of support from the therapist. ("Do you think I'm mentally ill, doctor, from the way I'm talking?" "Is this normal with a person?" "Did you come across anybody who had this, doctor? (Yes) Have you helped them?") Such solicitation was not evident in the first-session transcripts of differentiated patients.

---

[16] This behavior reflects strong open feelings of shame; such feelings seem prevalent among undifferentiated patients and less common (at least in open form) among differentiated ones. It may very well be that the "teeth pulling" in the first kind of undifferentiated patient may also have strong feelings of shame as one of its bases.

We are now devising ways of pursuing these observations systematically. First, we are attempting to establish criteria for identifying meaningful units of patient-therapist interaction. Second, we are working out methods of evaluating such units with regard to the variables I outlined.

A final aspect of psychotherapy we may consider is the achievement of change. The issue of change in therapy arises in two ways. First, will the characteristics of more developed or less developed differentiation themselves change in the course of therapy? Second, is the potential for change related to initial level of differentiation?

As for the first question, we are inclined to expect that the cluster of characteristics subsumed under differentiation, and the cognitive-style component in particular, are not likely to change with therapy. We expect this mainly because formal features of an individual's psychological make-up, of the kind represented by differentiation, are likely to show considerable stability as compared to content features.

There are already several lines of evidence from areas other than therapy, which support the expectation of marked relative stability of level of differentiation. Stability has been found even over long periods of time and in the face of strenuous attempts to bring about change.

First, longitudinal studies we have done demonstrate considerable relative stability during the growth years in the characteristics of differentiation we have thus far examined. In one study we followed a group of boys for 14 years, testing them at ages 10, 14, 17, and 24. Of the 30 boys first seen at age 10, we were still able to obtain 28 for testing at age 24. Considering group means, there was a change in the group as a whole, over the 14-year span, toward greater field independence and a more articulated body concept. At the same time, test-retest correlations were high, indicating that children who show more developed differentiation in the areas considered at one age, compared to their group, tend to have the same relative standing at other ages as well. Test-retest correlations for the rod-and-frame test [17] for various age combinations were: 10 versus 14, .75; 10 versus 17, .80; 10 versus 24, .66; 14 versus 17, .93; 14 versus 24, .86; and 17 versus 24, .93 (all significant). Test-retest correlations for figure-drawing sophistication-of-body-concept scores were equally high: 10 versus 14, .84; 10 versus 17, .79; 10 versus 24, .64; 14 versus 17, .67; 14 versus 24, .67; 17 versus 24, .68 (all significant). This evidence of stability is particularly impressive since a 14-year period was covered, and, moreover, a period when the child enters many new areas of life, when important new needs within himself are emerging, and when ways of coping with these new life circumstances and needs are being developed.

Comparable longitudinal data for adults (Bauman, 1951) again show striking stability in characteristics of differentiation over time. The test-retest correlations after a 3-year interval, for Bauman's group of young men, were .84 for rod-and-frame test scores and .86 for figure-drawing sophistication-of-body-concept scores (both significant).

Another kind of evidence of stability comes from studies which showed that the articulated-global cognitive style could not be changed by experimental intervention. The methods used in an attempt to bring about change have included drug administration (Karp, Witkin & Goodenough, 1965b; Pollack, Kahn, Karp & Fink, 1960), electroconvulsive shock (Pollack et al., 1960), stress due to anticipated heart surgery (Kraidman, 1959), special training (Gruen, 1955, Witkin, 1948), hypnosis.[18] Consistent with these results, Bauman (1951) found no difference in test-retest stability, over a 3-year period, between subjects who underwent important changes in life circumstances (as marriage, divorce, psychotherapy) and subjects who did not.[19]

---

[17] The rod-and-frame test is the only one of our tests of field dependence which seems to show no practice effect. It is therefore our best test for studies which require repeated testing, as this longitudinal study. The correlations are based on scores for the body-erect condition of the test, a shortened version of the test we now use.

[18] Personal communication from Rochelle M. Wexler.

[19] There is another kind of evidence, less direct in nature, which suggests that perceptual field de-

The nature of differentiation as a structural aspect of personality, and the stability of relative level of differentiation over time and with experimental intervention, lead us to think that the characteristics of differentiation may not change with psychotherapy. A check on this expectation remains to be carried out. It would be of particular interest to examine change with the most probing form of therapy now in use—psychoanalysis.

Let me turn now to the second question I raised about change in therapy—namely, whether more differentiated and less differentiated patients may differ in their potential for change. Even if it turns out that cognitive style and associated characteristics of differentiation themselves remain more or less stable during therapy, we may of course expect other important features of personality to change. A variety of modes of integration is possible at any level of differentiation. Reintegrations, leading to salient alterations in personality, may take place with a given level of differentiation fundamentally maintained. In this connection the children in our longitudinal studies, though presenting a stable picture with regard to relative level of differentiation, showed many important personality changes during the time we studied them. To answer the question of whether potential for change during therapy, of whatever kind, is at all a function of level of differentiation we must consider the dimension of "fixity-mobility" which relates to differentiation in a very complex way, and may actually cut across the differentiation dimension.

Our clinical observations strongly suggest that some persons who are field independent and show other characteristics of developed differentiation function consistently at a highly differentiated level, whereas, others vary more according to circumstances and inner state. The first kind of subject shows "fixity" of functioning, the second shows "mobility." The perceptual tasks we devised to assess cognitive style, by design, press the subject to perform analytically if he possibly can.[20] They therefore do not permit us to distinguish between these two kinds of subjects. Cognitive tests may of course be devised which leave it to the option of the subject to function at his maximum analytical capacity or not. Results of a study by Perez (1955), using a brightness constancy situation, suggest that with instructions which pushed them toward an analytical attitude subjects identified as field independent tended to perform in a consistently field-independent fashion; with instructions which encouraged a global attitude they varied more with respect to extent of field independence.

We may venture the prediction that patients who are relatively differentiated (and in this sense have developed resources available to them), and who at the same time show "mobility," would be the best candidates for change. On the other hand, undifferentiated

---

pendence may be under some degree of genetic control. With the ultimate aim of studying neurophysiological correlates of this perceptual dimension, and the effect of radical variations in early sensory and social experience upon its development, we have undertaken studies of field dependence in monkeys. Preliminary results with an embedded-figures type of task devised for monkeys show that at least some of these animals are quite adept at even difficult perceptual disembedding tasks (Rosenblum, Witkin, Kaufman, & Brosgole, 1965). This finding is perhaps not surprising when we consider that disembedding may be a perceptual function of great adaptive significance. Under feral conditions survival may depend on the ability to separate quickly from the complex visual and auditory fields in which the animals lives those sights and sounds which signify danger, food, etc. Functions which have such survival value, and which may therefore have been selected for, may well be under at least partial genetic control. In biological time man is not so very far away from his feral existence so that such genetic control may still exist in him. Stafford (1963) has recently presented some evidence which suggests that the capacity to perform tasks similar to those used to evaluate perceptual field dependence is under the influence of a recessive gene on the X chromosome (one of the sex chromosomes). Also relevant here is the observation that Turner women, who lack one of the two X chromosomes of normal women, are consistently and extremely field dependent as indicated by their very poor performance on the analytical triumvirate of Wechsler subtests (Shafer, 1962). These observations, if substantiated, are of particular interest in view of the repeated finding of significant sex differences in field dependence under a wide range of cultural conditions.

[20] For example, in the embedded-figures test the subject, on each trial, is told if his choice of the simple figure is incorrect, and he is required to continue his search for 5 minutes if he does not find it before then.

personalities, whose resources are limited and who patently cannot have great "mobility," would be the least likely candidates for change. It is further possible that patients who are intermediate in the range of measures of differentiation are in general more amenable to change than highly differentiated patients. This view, which remains to be checked, is based on the clinical impression that the latter group is likely to include relatively more patients with limited "mobility."[21]

Obviously, differentiation is only part of what must be considered in evaluating change in therapy. Should it turn out that cognitive style, for example, does not change, it would be of considerable interest. For a comprehensive view of the accomplishments of therapy we need to know what changes as well as what does not change. Changes in given characteristics may have a different meaning when occurring in the context of presence or absence of change in other characteristics.

I hope this account has helped suggest some of the ways in which the concept of differentiation, and the technique we have devised for the assessment of differentiation, may prove useful to the clinician in the problems he faces in his work in diagnosis and therapy.

---

[21] Numerous studies have shown that greater authoritarianism (as measured by the F scale) goes with greater field dependence (see, for example, Jackson, 1955; Linton, 1952; Pollack et al., 1960; Rudin & Stagner, 1958). An unpublished study by Mednick suggests, however, that the relation may be nonlinear; both field-dependent and field-independent persons seem to score higher for authoritarianism than intermediates. In view of the personal characteristics associated with F scale scores, it seems reasonable to consider these scores as rough indicators of "fixity" or "mobility."

REFERENCES

Alexander, F. *Psychosomatic medicine.* New York: Norton, 1950.

Andrieux, C. Contribution a l'etude des differences entre hommes et femmes dans la perception spatiale. *L'annee Psycholoqique,* 1955, **55**, 41–60.

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science,* 1953, **118**, 273.

Axelrod, S., & Cohen, L. D. Senescence and embedded-figures performance in vision and touch. *Perceptual & Motor Skills,* 1961, **12**, 283–288.

Bailey, W., Hustmyer, F., & Kristofferson, A. Alcoholism, brain damage and perceptual dependence. *Quarterly Journal of studies on Alcohol,* 1961, **22**, 387–393.

Bauman, G. The stability of the individual's mode of perception, and of perception-personality relationships. Unpublished doctoral dissertation, New York University, 1951.

Bennett, D. H. Perception of the upright in relation to body image. *Journal of Mental Science,* 1956, **102**, 487–506.

Bertini, M. Il tratto difensivo dell'isolamento nella sua determinazione dinamica e strutturale. *Contributi dell' Istituto di Psicologica,* 1961, Serie XXV. (Milano: Univer. Cattolica)

Bertini, M., Lewis, Helen B., & Witkin, H. A. Some preliminary observations with an experimental procedure for the study of hypnagogic and related phenomena. *Archivio di psicologia, neurologia e psichiatria,* 1964, **xxv**(6), 495–534.

Bound, Mae M. A study of the relationship between Witkin's indices of field dependency and Eysenck's indices of neuroticism. Unpublished doctoral dissertation, Purdue University, 1951.

Bryant, A. R. An investigation of process-reactive schizophrenia with relation to perception of visual space. Unpublished doctoral dissertation, University of Utah, 1961.

Cancro, R. A comparison of process and reactive schizophrenia. Unpublished doctoral dissertation, State University of New York Downstate Medical Center, 1962.

Chateau, J. Le test de structuration spatiale TIB. I. S. *Le travail humain,* 1959, **22**, 281–297.

Cohen, J. The factorial structure of the WAIS between early adulthood and old age. *Journal of Consulting Psychology,* 1957, **21**, 283–290.

Cohen, J. The factorial structure of the WISC at ages 7–6, 10–6, and 13–6. *Journal of Consulting Psychology,* 1959, **23**, 285–299.

Crutchfield, R. S., Woodworth, D. G., & Albrecht, Ruth E. Perceptual performance and the effective person. Lackland AFB, Texas, *USAF WADC tech. Note,* 1958, No. 58–60.

Dawson, J. L. M. Psychological effects of social change in a West African community. Unpublished doctoral dissertation, University of Oxford, 1963.

Dement, W., & Kleitmen, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346.

Epstein, L. The relationship of certain aspects of the body image to the perception of the upright. Unpublished doctoral dissertation, New York University, 1957.

Fishbein, G. M. Perceptual modes and asthmatic symptoms: An application of Witkin's hypothesis. *Journal of Consulting Psychology,* 1963, **27**, 54–58.

Franks, C. M. Differences déterminées par le personalité dans la perception visuelle de la verticalité. *Revue de Psychologique Appliques,* 1956, **6,** 235–246.

Freedman, N. The process of symptom modification in psychopharmacological therapy. Paper presented at Department of Psychiatry meeting, State University of New York Downstate Medical Center, Brooklyn, New York, 1962.

Goodenough, D. R., & Karp, S. A. Field dependence and intellectual functioning. *Journal of Abnormal and Social Psychology,* 1961, **63,** 241–246.

Gordon, B. An experimental study of dependence-independence in a social and a laboratory setting. Unpublished doctoral dissertation, University of Southern California, 1953.

Gruen, A. The relation of dancing experience and personality to perception. *Psychological Monographs,* 1955, **69** (14, Whole No. 399).

Holt, R. R., & Goldberger, L. Personological correlates of reactions to perceptual isolation. *WADC tech. Rep.,* 1959, No. 59–735.

Jackson, D. N. Stability in resistance to field forces. Unpublished doctoral dissertation, Purdue University, 1955.

Jannucci, Gloria I. Size constancy in schizophrenia: A study of subgroup differences. Unpublished doctoral dissertation, Rutgers State University, 1964.

Karp, S. A. Field dependence and overcoming embeddedness. *Journal of Consulting Psychology,* 1963, **27,** 294–302.

Karp, S. A., Poster, Dorothy, & Goodman, A. Differentiation in alcoholic women. *Journal of Personality,* 1963, **31**(3), 386–393.

Karp, S. A., Witkin, H. A., & Goodenough, D. R. Alcoholism and psychological differentiation: Effect of achievement of sobriety on field dependence. *Quarterly Journal of Studies on Alcohol,* 1965, in press. (a)

Karp, S. A., Witkin, H. A., & Goodenough, D. R. Alcoholism and psycholgical differentiation: The effect of alcohol on field dependence. *Journal of Abnormal Psychology,* 1965, **70,** 262–265.

Konstadt, Norma, & Forman, Elaine. Field dependence and external directedness. *Journal of Personality & Social Psychology,* 1965, **1,** 490–493.

Kraidman, Emma. Developmental analysis of conceptual and perceptual functioning under stress and non-stress conditions. Unpublished doctoral dissertation, Clark University, 1959.

Linton, Harriet B. Relations between mode of perception and tendency to conform. Unpublished doctoral dissertation, Yale University, 1952.

Linton, Harriet B. Dependence on external influence: Correlates in perception, attitudes, and judgement. *Journal of Abnormal & Social Psychology,* 1955, **51,** 502–507.

Messick, S., & Damarin, F. Cognitive styles and memory for faces. *Journal of Abnormal & Social Psychology,* 1964, **69,** 313–318.

Pardes, H., & Karp, S. A. Field dependence in obese women. *Psychosomatic Medicine,* 1965, **27,** 238–244.

Perez, P. Experimental instructions and stimulus content as variables in the size constancy perception of schizophrenics and normals. Unpublished doctoral dissertation, New York University, 1955.

Pollack, M., Kahn, R. L., Karp, E., & Fink, M. Individual differences in the perception of the upright in hospitalized psychiatric patients. Paper read at Eastern Psychological Association, New York, 1960.

Pollack, I. W., & Kiev, A. Spatial orientation and psychotherapy: An experimental study of perception. *Journal of Nervous & Mental Disease,* 1963, **137**(1), 93–97.

Powell, B. J. A study of the perceptual field approach of normal subjects and schizophrenic patients under conditions of an oversize stimulus. Unpublished doctoral dissertation, Washington University, 1964.

Rosenblum, L. A., Witkin, H. A., Kaufman, I. C., & Brosgole, L. Perceptual disembedding in monkeys: A note on method and preliminary findings. *Perceptual & Motor Skills,* 1965, **20,** 729–736.

Rudin, S. A., & Stagner, R. Figure-ground phenomena in the perception of physical and social stimuli. *Journal of Psychology,* 1958, **45,** 213–225.

Schafer, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

Schonbar, Rosalea, A. Some dimensions of sensitivity of recallers and nonrecallers of dreams. In (Chm.), Dream research and theory. Symposium presented at the Postgraduate Center for Mental Health, March 14, 1964.

Shafer, J. W. A specific cognitive deficit observed in gonadal aplasia (Turner's Syndrome). *Journal of Clinical Psychology,* 1962, **18**(4), 403–406.

Silverman, A. J., Cohen, S. I., Shmavonian, B. M., & Greenberg, G. Psychophysical investigations in sensory deprivation: The body-field dimension. *Psychosomatic Medicine,* 1961, **23,** 48–61.

Soll, J. The effect of frustration on functional cardiac disorder as related to field orientation. Unpublished doctoral dissertation, Adelphi University, 1963.

Stafford, A. E. An investigation in parent-child test scores for evidence of hereditary components. Unpublished doctoral dissertation, Princeton University, 1963.

Taylor, J. M. A comparison of delusional and hallucinatory individuals using field dependency as a measure. Unpublished doctoral dissertation, Purdue University, 1956.

Winestine, Muriel C. Twinship and psychological differentiation. Unpublished doctoral dissertation, New York University, 1964.

Wit, O. C. Sex differences in perception. Unpublished master's thesis, University of Utrecht, 1955.

Witkin, H. A. The effect of training of structural aids on performance in three tests of space orientation. *CAA Div. Res. Rep.* 1948, No. 80.

Witkin, H. A. Cognitive patterning in congenitally blind children. Presented at a symposium on

"Cognitive Structure and Personality," American Psychological Association, Chicago, Ill., 1965.

WITKIN, H. A., DYK, RUTH B., FATERSON, HANNA F., GOODENOUGH, D. R., & KARP, S. A. *Psychological Differentiation,* New York: Wiley, 1962.

WITKIN, H. A., FATERSON, HANNA F., GOODENOUGH, D. R., & BIRNBAUM, JUDITH. Cognitive patterning in high grade mentally retarded boys. *Child Development,* in press.

WITKIN, H. A., KARP, S. A., & GOODENOUGH, D. R. Dependence in alcoholics. *Quarterly Journal of Studies on Alcohol,* 1959, **20**, 493–504.

WITKIN, H. A., & LEWIS, HELEN, B. The relation of experimentally induced pre-sleep experiences to dreams: A report on method and preliminary findings. *Journal of the American Psychoanalytic Association,* in press.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M. MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception,* New York: Harper, 1954.

ZUKMANN, L. Hysteric compulsive factors in perceptual organization. Unpublished doctoral dissertation, New School for Social Research, 1957.

(Received October 20, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 337–342

RESIDUAL PSYCHOPATHOLOGY IN RELEASED PSYCHIATRIC PATIENTS AND ITS RELATION TO READMISSION[1]

ROY JOHNSTON AND BENJAMIN F. McNEAL
Veterans Administration Hospital, Canandaigua, New York

The degree and type of mental pathology reflected in MMPI scores was surveyed in 147 hospitalized psychiatric patients upon admission and release. Both statistical and clinical assessments of the profiles were made with analyses related to changes during hospitalization and to comparison with normal groups. Results vary depending on method of assessment. A substantial group of profile-pairs are judged as showing improvement at release (75%), while the remainder (25%) show more illness. Statistical analysis shows significant group improvement but exit profiles do not resemble those of normals. Psychotic profile types do not change to resemble neurotic ones with hospital treatment. MMPI change scores seem to anticipate rehospitalization better than evaluation of the patients' exit profile alone.

Hospital release, because of its objectivity, is frequently used as a measure of success in treating psychiatric patients. Considerable pride in the progress of hospital treatment is based on achieving higher release rates than in recent decades, and higher turnover rates and shorter hospital stays are being used to document the success of the tranquilizing drugs and to justify their expense. Hospitals compete with each other in this regard and semiautonomous units within hospitals calculate their success by rates of release. An objective inspection of the mental condition of patients who leave the psychiatric hospital appears necessary to evaluate the validity of this measure.

We are concerned with the following questions: How do the quality and quantity of emotional disturbance change during hospitalization? How closely do measures of patients' emotional control upon hospital release resemble those of normals? Are these related to patients' length of stay outside before next admission? Are they related to total amount of time spent out of the hospital during a 3-year follow-up period?

METHOD

The Minnesota Multiphasic Personality Inventory (MMPI) scores were used as criteria of mental pathology. Psychiatric patients were administered this test upon admission to our hospital as part of another research project. For the present study, the test was also administered to 160 consecutive patients leaving the hospital. The tests were scored several weeks later; thus they did not influence staff decision to release patients. This represented 90% of the available sample and excluded only patients who refused testing, left the hospital without permission, or were missed since they left by way of conversion of a short visit home into a trial visit. The average age of these patients was 40.5 years, 96% were white, and 80% were diagnosed psychotic (nearly all schizophrenia); the remaining 20% carried neurotic or character disorder diagnoses. Forty-six percent had been hospitalized for 3 months or less, 43% for 4 to 12 months, and 11% for over a year. Of the 160 patients tested at the time of release, 147 had been tested at the time of admission. Admission tests were administered within 2 weeks after admission. Exit tests were obtained within 2 weeks of the patient's trial visit or discharge from the hospital.

For statistical comparison, the average *K*-corrected *T* scores of the validity and conventional clinical scales were computed for the admission and exit samples and differences evaluated by critical ratio tests. The average *T* score values of the exit sample were also compared with those of two normal control groups. Three measures of type of pathology were derived from the MMPI. (*a*) The high point codes were tabulated and admission-exit changes noted. (*b*) The Meehl-Dahlstrom PNI rules (Meehl & Dahlstrom, 1960) were applied to both admission and exit profiles. (*c*) The score profiles were evaluated clinically. Seven psychologists blindly selected the better of each admission-exit profile pair and nine psychologists made 3-point mental-health-scale ratings of each exit profile. In addition all exit records were scored on the Forsyth Recovery Scale (1960).

RESULTS

As a group our patients showed significant improvement on most of the MMPI scales

---

[1] The authors wish to thank Leonard Coleman for his assistance in scoring and interpreting the numerous MMPI protocols.

TABLE 1

AVERAGE T SCORE MEANS, STANDARD DEVIATIONS, AND DIFFERENCES BETWEEN ADMISSION AND EXIT MMPI SCORES FOR 147 INPATIENTS AND DIFFERENCES BETWEEN THEIR EXIT PROFILES AND THOSE OF TWO NORMAL CONTROL SAMPLES

MMPI scale	Patient admission versus exit						Patient exit versus normals			
	Admission		Exit				Sample A		Sample B	
	M	*SD*	*M*	*SD*	Difference	*t*	Difference	*t*	Difference	*t*
L[a]	53.2	11.5	54.9	9.3	−1.7	1.04			4.9	5.23**
F	60.9	8.4	54.6	10.6	6.3	5.89**			4.6	3.65**
K[a]	55.4	11.0	61.3	10.9	−5.9	4.80**			10.3	9.45**
Hs	62.5	8.2	58.0	15.4	4.5	3.26**	3.3	1.81	8.0	11.87**
D	68.9	13.4	60.3	16.1	8.8	5.33**	5.6	2.71**	10.3	8.51**
Hy	63.8	9.4	60.1	11.4	3.7	3.16**	3.6	2.25*	10.1	11.01**
Pd	68.6	9.8	65.3	14.4	3.3	2.39*	12.0	6.32**	15.3	15.77**
Mf	60.2	11.0	57.5	10.0	2.7	2.29*	2.0	1.18	7.5	7.08**
Pa	60.8	8.9	55.9	13.5	4.9	3.80**	2.1	1.34	5.9	6.48**
Pt	66.1	11.2	60.0	13.3	6.1	4.45**	7.2	3.91**	10.0	9.26**
Sc	68.1	11.0	60.9	18.0	7.2	4.31**	8.8	4.68**	10.9	10.28**
Ma	61.6	13.7	58.0	14.2	3.6	2.29*	7.3	3.23**	8.0	6.45**

Note.—Sample A: 38 normal males (Sopchak, 1958); Sample B: Minnesota Normal males (Dahlstrom & Welsh, 1960).
[a] Exit scores higher than on admission: on all other scales the exit scores were lower.
$^{*}p < .05$.
$^{**}p < .01$.

under consideration. All mean T score clinical scale values were less elevated upon exit. Comparisons between the exit profiles and scores of a normal male sample (Sopchak, 1958), however, show significantly higher scale scores among the patient group in six of the nine clinical scales (*D*, *Hy*, *Pd*, *Pt*, *Sc*, and *Ma*). Comparison with the Minnesota male norms (Dahlstrom & Welsh, 1960, p. 48) show our patient group to be significantly more deviant on all clinical scales. While improvement in a relative sense is demonstrated, the patient group still differs markedly in some aspects from the nonpsychiatric controls and from the general population. These data are summarized in Table 1.

Elevations of the MMPI standard score values give additional normative data on degree of psychopathology. Presence of one primed scale (T score ≥ 70) in the profile and the average T score value of the nine clinical scales have been used as measures of severity of mental illness (Modlin, 1947). Scale elevations below 70 are considered within normal limits. A statistical comparison between the admission and exit MMPI profiles revealed a significant reduction of elevated profiles ($\chi^2 = 42.72$, $p < .001$). Comparison data are presented in Table 2.

The Meehl-Dahlstrom PNI rules, reflecting another measure of mental pathology, were applied to the admission and exit protocols. No change in the PNI category was found among 66 (45%) of the patients during their hospital stay. Specifically for 29 patients both test profiles were scored as Psychotic[2] (P), 9 as Neurotic (N), 17 as Indeterminate (I), and 11 as Nonapplicable (NA). Four patients scored as P achieved N ratings upon exit and 5 patients initially scored N became P. Even making the plausible assumption that, for our patients, the I ratings reflect psychopathology between the P and N ratings and excluding the NA group from consideration, we find that 24 patients achieved a "better" rating on exit while 27 achieved a "poorer" rating. As a group the sample showed no significant change in the P, N, I classes from admission to exit ($\chi^2 = 7.21$, $p < .10$). Comparison by percentage is presented in Table 3.

The designation NA indicates that the protocol is considered invalid for the PNI scoring system because of an excessive elevation of one or more of the MMPI validity

---

[2] The percentage of P and N profiles varies with treatment setting. The figure of 30% Indeterminate profiles reported by Meehl and Dahlstrom (1960) compares well with our findings of 27% and 30%.

TABLE 2

NUMBER OF MMPI PROFILES HAVING ONE OR MORE CLINICAL SCALE SCORES WITHIN VARIOUS T SCORE RANGES

T score	Admission	Exit
<70	31	63
70–79	33	42
80–89	40	25
≥90	42	16

scales. Such profiles are often found among acutely disturbed or uncooperative psychotic patients. Elevation of the *F* scale accounted for 20 of the 32 admission NA profiles, while elevated ? and L scales accounted for the remaining 12. Upon exit a reverse trend was found: elevated ? and *L* scales accounted for 14 of the 19 NA exit profiles. This suggests that within the NA subgroup there was less endorsement of infrequent or obviously bizarre items on exit than on admission, and there was a tendency toward increased endorsement of items reflecting falsification and undue defensiveness; the latter presumably an effort to appear symptom free.

Clinical Evaluation

Since clinical usage of the MMPI is concerned with the individual patient, some attention must be paid to psychologists' judgments of each profile in evaluating level of psychopathology. Unlabeled pairs of admission and exit MMPI profiles were drawn on MMPI profile sheets, one sheet for each patient. The 147 admission-exit graphs were then submitted for clinical judgment. Six psychologists and one psychology technician familiar with the MMPI selected the better, that is, the more healthy of the two. A majority of judges (4 of 7) chose the exit profiles as more healthy in 75% of the cases (Range = 72.8%–79.2%).

TABLE 3

PERCENTAGE CLASSIFICATION OF PATIENT SAMPLE BY MEEHL-DAHLSTROM PNI RULES[2]

PNI category	Admission	Exit
Psychotic	37	43
Neurotic	14	14
Indeterminate	27	30
Nonapplicable	22	13

Reliability of admission-exit comparisons between judges evaluated by phi-correlation coefficients ranged from .65 to .88, and averaged .75. The judges independently listed their own particular schema and criteria used in assessing the profiles. Their comments showed good agreement. Most considered elevation of the neurotic over the psychotic triad, lower *F*, lower *D*, higher *Ma*, higher *K*, and an overall drop in all clinical scales as indicators of improvement.

A purely mathematical check obtained by selecting the profile having the lower total *T* score value among the clinical scales as the more healthy, included 73.5% of the exit profiles. In such a numerical classification there should be a "no change" category to avoid undue influence of chance fluctuations. Arbitrarily, reduction of the sum of the *T* scores on the clinical scales by 20 points or more was considered improvement, increase by 20 or more as worsening, and values in between as representing no change. By these criteria, 62% of the group improved, 19% became worse, and 19% showed no change.

To approximate a practical clinical use of the MMPI, nine clinical psychologists were asked to assume they had each exit profile as the only bit of information available in making one of three alternative release recommendations at disposition staffing. The three rating categories were: Yes, patient may leave the hospital; No, patient should remain in the hospital; Yes, patient may leave but I have doubts and reservations about his later adjustment.

We recognize that psychologists working in a mental hospital may make a release recommendation even though the exit profile is not free of obvious psychopathology. Judges afterwards remarked that they rated all profiles as "Yes" except in cases where scale elevations strongly suggested that the patient was a danger to himself or to others. The release recommendation data for the judges as a group are summarized in Table 4. Calculated by majority of opinion, that is, by endorsement of 5 of 9 judges, 64% of the profiles were rated Yes, 14% No, 8% With Reservations, and 14% no majority. Thirty-six percent were rated as either No or With Reservations. Unanimous agreements were

TABLE 4

MENTAL HOSPITAL RELEASE RECOMMENDATION BY NINE PSYCHOLOGISTS FROM 160 MMPI EXIT PROFILES

Recommendation	Average	Percentage	Range
Yes	98.6	61.6	74–129
No	28.0	17.4	9-45
Reservation	33.4	21.0	14–56

less prevalent: 25% achieved a Yes rating and 4% a No rating. While the range of judgments for each category was large, phi-correlation coefficients between pairs of judges ranged from .59 to .81, averaging .72. Maximum phi coefficients possible averaged .74 suggesting that the disagreements were on a few ambiguous profiles.

Relationship to Hospital Readmission

The measures of mental adjustment used were believed to be related to two criteria of mental health following hospital release: (*a*) time (in months) before readmission, and (*b*) total time out of the hospital during the 3-year follow-up period. Complete follow-up information was obtained on all service-connected veterans by correspondence with the Veterans Administration Regional Offices and Veterans Administration Hospitals. Information on non-service-connected veterans was obtained from the same sources in all except 16 cases. Next of kin replied to our direct inquiries in 12 of these cases leaving 4 cases unaccounted for. Within 3 months after release, 23 patients or 16% of the total sample returned to a psychiatric hospital, within 6 months a total of 27% had returned, and within 1 year 37% had returned. Two years after their release a total of 51% of the total sample had returned and 56% had returned within 3 years. Sixty-five patients or 44% remained out continuously beyond the 3-year follow-up period.

TABLE 5

RELATIONSHIP BETWEEN VARIOUS PRE-EXIT PREDICTOR VARIABLES AND MMPI MEASURES TO REHOSPITALIZATION

Predictor variable	Time out before next admission	Total hospital-free time in 3 years
Clinical judgment		
Admission-exit comparison	$X^2 = 5.50$ $p < .02$	$X^2 = 8.16$ $p < .02$
Release recommendation at exit	$X^2 = 8.18$ $p < .02$	*ns*
Peak exit score of 6, 7, 8, or 9	*ns*	*ns*
Primed codes at exit	*ns*	*ns*
Diagnosis on admission	*ns*	*ns*
Length of stay	*ns*	$X^2 = 9.04$ $p < .05$

Patients with profiles judged as improved remained out an average of 23.3 months before readmission, while those with worse release profiles remained out 16.8 months. Similarly the improved profile group spent an average of 30.3 months out of the hospital during the 3-year follow-up period, while the worse group averaged 25.7 months. While MMPI admission-exit change scores are related significantly (by chi-square test) to rehospitalization, cutting scores for individual prediction could not be established. Although not statistically significant, the trend of the data also suggests a relationship between the group judgment and the total hospital-free time in the follow-up period. Good profiles indicate a long stay out, but poor profiles are not predictive of a quick readmission. The measures tested relating to rehospitalization are presented in Table 5.

In his doctoral dissertation, Forsyth (1960) produced a 22-item Recovery Scale of MMPI items which differentiated patients receiving adaptive versus maladaptive ratings on posthospital adjustment. There was no significant relationship between exit scores on Forsyth's scale and readmission data or total hospital-free time in our follow-up sample. The Recovery scale was however related to the group judgment of mental health upon exit ($\chi^2 = 22.55, p < .001$).

DISCUSSION

It may be considered an asset in this survey that no patients capable of responding to the MMPI were excluded because of any diagnostic or demographic reason. Results obtained are not related to special treatment programs or modalities but rather focus upon the impact of the total hospital regimen on an unselected and representative sample of our psychiatric patient population.

As an objective and convenient measure of mental disturbance, the MMPI was chosen as appropriate for assessing the psychiatric progress of patients released from a mental hospital. We realize that hospital patients who are well controlled may withhold gross mental symptoms on psychological tests (Grayson & Olinger, 1957). Thus, if error occurs here in evaluating extent of psychopathology, it will result in minimizing or hiding, rather than accentuating, the degree of disturbance among our sample. Some bias probably exists in our test popuation since a higher proportion of severely disturbed patients on admission and those considered more "sick" on release might be found in the approximately 10% of the patients from whom tests could not be obtained.

Our results suggest that release from a mental hospital is only generally indicative of improvement in terms of MMPI performance. About one-fourth of those released appear more ill on exit than on admission. The less healthy patients on exit tend to return sooner to the hospital and spend more of the next 3 years of their life in mental hospitals than do those patients who show improvement. Hospital release may be justified by criteria other than the absence, or reduction, of symptoms of mental pathology. Release of the "sicker" group reflects the current philosophy of rapid patient turnover as a therapeutic manipulation for the patient's benefit designed to avoid bad effects of prolonged institutionalization.

While as a group, the sample showed less pathology upon exit, they still differ as a group from the normal controls. Similar findings were presented by Kaufman (1950) in evaluating outcome of 22 counseling cases. His patient group was still differentiated from his control group on 6 of 9 MMPI clinical scales on posttesting. Schofield (1950) examined MMPI changes in various diagnostic groups of women after ECT and psychotherapy and found differences between groups in degree of improvement reflected in MMPI scores. Psychotics without ECT showed no changes while those receiving ECT had significant decreases in the *Hs, D, Hy, Pd, Pt,* and *Sc* scores. Schofield and Briggs (1958) report that 61% of the MMPI profiles of their diagnostically heterogeneous sample of hospitalized patients were judged as improved after treatment, even fewer than our finding of 75% judged improved.

MMPI scales reflecting acting-out and display of obvious and severe mental symptoms (*F, Pd, Pa, Sc*) show the greatest change with hospital stay but also still differentiate the released patients from normals. Similarly, *F, Pd, Pa,* and *Sc* scales were found most useful by Haertzen (1952) in differentiating discharged from hospitalized patients and *F* and *Sc* correlated best with adjustment criteria in a 1-year follow-up (Marks, Stauffacher, & Lyle, 1963).

The lack of marked group changes in the Meehl-Dahlstrom PNI rules during hospitalization was disappointing but corroborated findings of Simovich (1961). We may speculate that the deviant MMPI profiles of patients with serious mental disorders are not necessarily reactions to emotional episodes but represent the basic responses for these members of our chronic mental hospital population, either in or out of the institution. Release from a hospital, almost by definition implies some behavioral changes and hopefully some signs of better emotional and intellectual control. Some discrepancy apparently remains between observed clinical behavior and endorsement of statements reflecting mental pathology in the MMPI. There was no evidence in this study that psychotic patients when in remission, resort to neurotic defense patterns and thus resemble neurotics psychometrically. To the contrary, by PNI rules, nine more profiles were classified as Psychotic on hospital release than on admission.

The more disturbed exit profiles in terms of primed codes, high point code, or the diagnosis of psychosis did not forecast patients' post-release history over the 3-year follow-up period. Similar difficulty in prediction was demonstrated by Marks et al. (1963) who examined 111 predictor variables in a one-year follow-up study. Forsyth and Fairweather (1961) surveyed 66 variables including MMPI scores and concluded that psychological test scores obtained at time of release simply are not predictive of post-hospital behavior. For such prediction, we found that test *change* scores between admission and exit are more useful than exit evaluation alone. Apparently, we need to know each patient's particular mode of functioning and psychic baseline to assess his relative improvement or worsening. The successful use of longitudinal data in predicting remission of schizophrenia has been demonstrated by Vaillant (1964). His scoring system characterized as "mental invalids" ex-patients in the community who failed to achieve social remission as well as those who remain in the hospital. In addressing the problem of forecasting the adjustment of people following therapeutic intervention, we feel that the focus of investigation should be on *change* rather than on either initial or on final mental status.

Our findings indicate that a certain group of psychiatric hospital patients are released who still complain of considerable discomfort and who freely reveal many severe symptoms of mental illness. Viewed another way we may congratulate ourselves that a substantial group of patients released show significant improvement following care and treatment. While the prompt release of patients who are over the acute phase of their disturbance follows current psychiatric practice, rehospitalization of the more fragile of this group is a symptom of the community's inability as yet to cope with severe psychiatric problems. The MMPI may be useful in pointing out potential problems where more psychiatric attention and community resources would be necessary to effect a successful and lengthy adjustment outside the hospital.

REFERENCES

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook.* Minneapolis: Univer. Minnesota Press, 1960.

FORSYTH, R. P. MMPI and demographic correlates of post-hospital adjustment in neuropsychiatric patients. (Doctoral dissertation, University of North Carolina) Ann Arbor, Mich.: University Microfilms, 1960, No. 378.756 F 735m.

FORSYTH, R. P., & FAIRWEATHER, G. W. Psychotherapeutic and other hospital treatment criteria: The dilemma. *Journal of Abnormal and Social Psychology,* 1961, **62**, 598–604.

GRAYSON, H. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *Journal of Consulting Psychology,* 1957, **21**, 73–77.

HAERTZEN, C. A. The value of the MMPI as a predictor of discharge of mentally sick patients at the Rochester State Hospital of Minnesota. *Blocks and Blots,* 1952, **2**, 4–5.

KAUFMANN, P. Changes in the MMPI as a function of psychiatric therapy. *Journal of Consulting Psychology,* 1950, **14**, 458–464.

MARKS, J., STAUFFACHER, J. C., & LYLE, C. Predicting outcome in schizophrenia. *Journal of Abnormal and Social Psychology,* 1963, **66**, 117–127.

MEEHL, P. E., & DAHLSTROM, W. G. Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *Journal of Consulting Psychology,* 1960, **24**, 375–378.

MODLIN, H. C. A study of the MMPI in clinical practice with notes on the Cornell Index. *American Journal of Psychiatry,* 1947, **103**, 758–769.

SCHOFIELD, W. Changes in response to the MMPI following certain therapies. *Psychological Monographs,* 1950, **64**(5, Whole No. 311).

SCHOFIELD, W., & BRIGGS, P. R. Criteria of therapeutic responses in hospitalized psychiatric patients. *Journal of Clinical Psychology,* 1958, **14**, 227–232.

SIMOVICH, G. Test-retest changes during the course of hospitalization among some frequently occurring profiles. *Dissertation Abstracts,* 1961, **21**, 2787.

SOPCHAK, A. L. Spearman correlations between MMPI scores of college students and their parents. *Journal of Consulting Psychology,* 1958, **22**, 207–209.

VAILLANT, G. E. Prospective prediction of schizophrenic readmission. *Archives of General Psychiatry,* 1964, **11**, 509–518.

(Received January 14, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 343–349

FINGER ELECTROMYOGRAPHIC ACTIVITY DURING SLEEP:

ITS RELATION TO DREAMING IN DEAF AND NORMAL SUBJECTS[1]

JOHANN MARTIN STOYVA
University of California Medical Center, San Francisco

Observations carried out during the sleep of 13 deaf Ss showed that: (a) the rates of dream recall from rapid eye movement (REM) periods were similar to those for normal hearing Ss; (b) finger electromyographic (EMG) bursts outside of REM periods were not related to the recall of mental activity; (c) in both deaf and hearing Ss, REM periods showed a consistently accelerated rate of finger EMG activity in comparison with other stages of sleep; (d) contrary to expectations, rates of finger EMG activity for 10 normal hearing Ss were just as high as those of the deaf group. The implications of this finding for the motor theory of thinking were discussed.

In discussions of the motor theory of thinking, among the most frequently cited studies—along with those of Jacobson (1929, 1932)—have been Max's experiments on the deaf (1935, 1937). Max investigated the electromyographic (EMG) activity of deaf subjects during sleep and discovered a close relationship between the occurrence of muscle potential in the arm and finger muscles, and the subsequent recall of dreams. On 33 occasions when he awakened his subjects shortly after such bursts, Max obtained 30 instances of dream recall. But in 62 awakenings following periods of EMG quiescence, he obtained only 10 reports of dreams. There was, in other words, a strong association between finger EMG bursts and the recall of dreaming.

Of considerable interest in the light of recent research on sleep, was an ingenious speculation put forward by Max (1935). In examining those few occasions when his deaf subjects reported dreams after a quiescent EMG record, he surmised that "the dream content was primarily visual in nature, and the seat of activity may have been in the eye muscles which at the time were not in the electrical circuit [p. 481]." Had Max further pursued this remarkable conjecture, he might very well have demonstrated the association between dreaming and eye movements some two decades prior to the unexpected discovery by the physiologists Aserinsky and Kleitman (1953). Briefly, their finding was that bursts of eye movements occurred in a cyclical fashion during sleep. These bursts of rapid eye movements (REMs) regularly appeared during the periods of low-voltage fast waves characteristic of activated EEG (Stage 1 sleep). Awakenings from Stage 1 rapid eye movement periods (REM periods) produced a high frequency of dream recall; awakenings at other times (Stages 2, 3, and 4 sleep) yielded only a low incidence of recall. These observations were confirmed by Dement and Kleitman (1957b), who concluded that mental activity during sleep was confined to REM periods.

Because the work of Aserinsky and Kleitman has been amply confirmed and elaborated in subsequent studies (Goodenough, Shapiro, Holden, & Steinschriber, 1959; Kamiya, 1961; Rechtschaffen, Verdone, & Wheaton, 1963) further details will not be discussed in this paper. Of pertinence to the current investigation, however, was the study of Foulkes (1962), which provided strong evidence that recallable mental activity occurred *outside* of REM periods (non-REM recall) as well as *within* REM periods. Foulkes noted, though, that the non-REM reports differed from the REM reports in that they were more thoughtlike, were simpler and more closely related to everyday life, and often consisted only of static images.

---

[1] This research was supported partly through Grant M-6667(A) from the National Institute of Mental Health, United States Public Health Service, while the author was at Dalhousie University, and partly through a National Institute of Mental Health Postdoctoral Fellowship in the Interdisciplinary Training Program at the University of California Medical Center, United States Public Health Service Grant Number 5T1MH-7082. Grateful acknowledgement is made to William Noseworthy for his assistance in the research, and to Joe Kamiya, both for his valuable suggestions and for making research facilities available under Grant MH-05069-03 from the National Institute of Mental Health, United States Public Health Service.

The validity of non-REM reports, however, has been the focus of some controversy. It can be argued that this material represents recall from a preceding REM period, or that it is simply an awakening artifact. With REM recall, a close relationship has been demonstrated between dream reports and various physiological measures (Dement & Kleitman, 1957a; Kamiya, 1961; Roffwarg, Dement, Muzio, & Fisher, 1962). Efforts to demonstrate similar relationships in the case of non-REM recall have been notably unsuccessful. An exception was the recent work of Kamiya (personal communication) in which a positive association was found between non-REM recall and the presence of increased respiratory rate immediately prior to the moment of awakening.

It was felt that the case for non-REM mental activity would be further strengthened could additional evidence of physiological or behavioral correlates be found. Consequently, in the light of Max's work, the experimenter hypothesized that there might be a relationship between the occurrence of finger EMG bursts and the recall of non-REM mental activity in deaf subjects.

An additional point of interest in the current study concerned the *amount* of finger EMG shown by normal hearing subjects in comparison with the amount shown by deaf persons. Would there be a pronounced difference in favor of the deaf subjects? Such a difference was expected on the basis of Max's (1935) statement that "In control observations of 11 hearing subjects during 33 sleep periods (13 photographed, 20 observed), no instance of such action-current activity in either arms or legs was noted [p. 482]." Wolpert (1959, 1960), who more recently studied sleep EMGs of normal subjects, found definite evidence of arm and wrist activity in these individuals during REM periods. He surmised that this EMG activity had very likely escaped Max's notice since it was probably negligible in comparison with the extremely high rate of EMG activity shown by his deaf group.

Evidence of a higher rate of EMG activity on the part of deaf subjects would lend support to the motor theory of thinking—the idea that there are certain minimal motor activities closely associated with thinking processes. Max's work has, in fact, frequently been cited as constituting some of the most ingenious evidence in favor of such a theory. As Osgood (1953) stated in discussing Max's findings, "Since deaf-mutes 'talk' with their fingers, their mental activities should be reflected in more motor activity at this locus than with normal controls [p. 650]." Max's finding of a high rate of EMG activity in the deaf, as contrasted with a complete lack of such activity in hearing persons, could, therefore, readily be interpreted as confirming a motor theory of thinking. It was noted, however, that Max lacked any means of accurately quantifying activity—bursts were tallied either by photographing or by means of visual inspection. Consequently, the experimenter felt it would be of value to repeat Max's observations using the more refined instrumentation which had become available in the past quarter century.

In summary, the questions under investigation were these:

1. Does the EMG activity associated with finger movements in deaf subjects indicate the occurrence of non-REM mental activity?

2. What is the frequency and distribution of finger EMG activity during the sleep of deaf subjects? How does the amount of finger EMG activity shown by deaf subjects in the various stages of sleep compare with the amount shown by normal controls?

PART I

Method

Subjects. In the first phase of the study, seven deaf male subjects were recruited. Their ages ranged from 17 to 54 years with a mean of 31 years. The criteria of selection were: (*a*) that the deaf subject regularly communicated with other deaf persons using the manual alphabet, (*b*) that the subject's deafness be of long standing. Of the individuals chosen, three were congenitally deaf; four had become deaf prior to 2 years of age.

In this initial phase of the investigation only four hearing control subjects were employed. Their ages ranged from 20 to 28 years with a mean of 24 years. It was felt that this smaller sample of normal subjects would be sufficient to check Max's observation that normals show no EMG activity during sleep.

Laboratory procedure. Brain waves and eye movements were recorded using the method described by Dement and Kleitman (1957a). With each subject, the following attachments of the silver disc electrodes were made: (*a*) parietal lobe versus a left ear reference electrode, (*b*) occipital lobe versus a left ear reference electrode, (*c*) left eye versus right eye, for detection of lateral eye movements. Records were taken with a Grass Model III-A EEG.

In order to detect EMG arising from finger movements, two silver disc electrodes—each the size of a quarter—were attached to the subject's forearms. One electrode was placed 1 inch proximal to the crease of the wrist and medially; the other electrode was attached 2 inches distal to the elbow and over the extensor digitorum. If any electrode resistance exceeded 10,000 ohms, the electrode was reapplied. Note should be made that throughout this experiment it was discrete *bursts* of EMG activity which were being measured—not the sustained and rather regular background muscle tonus which may also be recorded with appropriate apparatus. Further, in this paper, the expression "EMG activity" always refers to bursts of EMG activity associated with finger movements.

Subjects slept in a darkened room and were awakened from three to six times in the course of a night. The experimenter, who monitored the subject's EEG from the adjoining machine room, awakened a subject by first switching on a soft light, and then, to ensure rapid awakenings, immediately shaking the side of the subject's bed in a brisk manner.

The subject was next handed a clipboard and asked to write out his recall on the attached form. On this form he was asked: (a) "Were you dreaming or thinking about anything *just before* I woke you up?"; and (*b*) "What was the *very last thing* that happened in your dream?"

In Part I of the study, subjects were run for a minimum of three nights each. These nights were not consecutive but were spaced out over several weeks.

Awakenings were made both within REM periods and outside them. Of the 119 awakenings, 68 of them, or 57%, took place outside of REM periods. Each of these non-REM awakenings was either:

1. An awakening closely following an EMG burst. In this condition, the experimenter would wake the subject up as quickly as possible after an EMG burst.

2. An awakening after the EMG leads had been quiet for some time. Under this condition, the experimenter tried to waken the subject after a long period of EMG quiescence.

In both Conditions "1" and "2," no awakenings were made while gross body movements were occurring, or for a 5-minute period subsequent to any body movement (detected by the appearance of muscle tension on the EEG channels). This was to minimize the possibility that some non-REM reports might be recall from partial awakening occurring during the body movement. Similarly, non-REM awakenings were not made until 20 minutes had elapsed since the end of the last REM period. This was to reduce the possibility that some of the non-REM material might simply be recall from the preceding REM period.

Results

Table 1 shows the percentage of dream recall from both REM and non-REM awakenings. Only recall of specific content was included in this table. Thus, if a subject wrote, "Yes, I was dreaming," but did not report any particular item of content, this response was classified as "no recall." For the seven subjects, mean recall from REM awakenings

TABLE 1

Dream Recall by Deaf Subjects from REM and Non-REM Sleep

Subject	REM				Non-REM					
					EMG			Non-EMG		
	Nights	Awakenings	Recall	Recall Percentage	Awakenings	Recall	Recall Percentage	Awakenings	Recall	Recall Percentage
FF	7	12	10	83	4	0	0	4	1	25
JM	6	13	6	46	8	2	25	5	1	20
KF	5	4	4	100	7	5	71	5	4	80
LD	4	6	5	83	6	2	33	5	2	40
JB	4	10	0	0	4	0	0	5	0	0
BF	4	3	3	100	4	3	75	4	1	25
CD	3	3	3	100	3	2	66	4	3	75
Mean Recall				73			39			38

Note.—$N = 7$.

was 73%. Mean recall from non-REM awakenings was 38%.

In the analysis of non-REM recall, awakenings were classified as either EMG or non-EMG awakenings. In the EMG category, the mean elapsed time between the last EMG burst and the moment of awakening was 24 seconds. Awakenings were classified in the non-EMG category if more than 2 minutes had elapsed since the last EMG burst. This relatively short cutoff period was made necessary by the high frequency of EMG bursts in some subjects. In the non-EMG category, the mean time between the last EMG burst and the moment of awakening was 6.0 minutes.

As may be seen in Table 1, the percentage of non-REM recall was virtually identical under the two conditions. In other words, Table 1 shows that there was no relationship between finger EMG activity and the recall of dreaming from non-REM sleep. Moreover there was no trend towards increased recall with shorter time lapses between the occurrence of the EMG bursts and the moment of awakening. Recall was just as likely following a 4-minute interval as it was after a 10-second interval.

In scoring EMG bursts, the criteria were generally the same as those described in Part II of the study with the exception that the amplitude of a burst had to equal or exceed 50 microvolts in order to reach criterion size. Table 2 shows EMG activity in terms of number of bursts per minute occurring both within and outside REM periods. For each subject, the figures given represent the mean rates for his first three nights. Even though some subjects had been run for considerably more than this minimum amount (see Table 1), the EMG analysis was confined to the first three nights of each subject in order to increase comparability of the data.

The four subheadings of Table 2 are arranged in the usual sequence in which REM and non-REM sleep occurs. "Spindles + 15" refers to the first 15 minutes of sleep. "Stages 2, 3, 4" indicates the time between the end of the first 15 minutes of sleep and a point 15 minutes prior to the onset of Stage 1 REMs. The expression "15 Pre-REM" refers to a period extending from 15 minutes prior to the onset of Stage 1, to the actual onset of Stage 1. And "REM period" indicates Stage 1 REM periods.

TABLE 2

FINGER EMG RATE IN DEAF VERSUS HEARING SUBJECTS: PART 1

Phase of sleep cycle	Deaf		Hearing	
	M	Range	*M*	Range
Spindles + 15	.38	.09–.65	.30	.14–.52
Stages 2, 3, 4	.21	.07–.52	.20	.08–.37
15 Pre-REM	.25	.06–.47	.26	.18–.30
REM period	1.23	.64–2.00	1.23	.89–1.80
Subject *N*		7		4
Night *N*		21		12

Note.—Calculated as bursts per minute.

As Table 2 shows, the mean EMG burst rates for the deaf and hearing subjects both during and outside of REM periods were practically identical. Only for the first 15 minutes of sleep was the mean for the deaf subjects somewhat higher.

PART II

In view of the unexpected finding of virtually identical finger EMG rates for deaf and hearing subjects, it was decided to replicate this part of the experiment using more sensitive instrumentation and scoring techniques. Each of the new subjects was given one night of uninterrupted sleep—in distinction to the practice of Part I, where several awakenings were made on each night.

Subjects. Six deaf and six hearing subjects were obtained. Four of the deaf individuals were congenitally deaf; two had become deaf at 3 months of age. Each of the deaf subjects used the manual alphabet when communicating with other deaf persons. Ages in the deaf group ranged from 25 to 38 years with a mean of 33 years. Ages of the normals ranged from 20 to 27 years with a mean of 23 years. All 12 subjects were male.

Instrumentation. EEG recordings were taken with an Offner Type-R, and EMG was recorded using two Offner Integrator couplers, Type 9852. Gains on the two EMG channels were adjusted to maximum sensitivity.

Electrodes were placed so as to replicate as closely as possible the positions described by Max (1935). Two silver electrodes—each the size of a quarter—were attached to each forearm. One electrode was placed 2 inches proximal to the distal crease of the wrist and medially on the forearm. The second electrode was placed 5 inches promixal to the distal crease of the wrist and also medially. Both these electrodes lay over the flexor digitorum sublimus. All resistances were 10,000 ohms or less with a mean reading of 5,000 ohms. Proper operation of the EMG electrodes was checked by instructing the subject to make very tiny finger movements and observing the corresponding EMG activity.

Scoring. An EMG burst was defined as any deflection equal to or exceeding 5 millimeters; that is, 20 microvolts or over. (At an EMG frequency of 70 cycles per second, a 1 millimeter pen excursion equalled 4 microvolts.) The EMG score was the duration of a burst in seconds, with any fraction of a second counted as a full second. In order to prevent confounding the data with gross body movements, EMG bursts were only scored if they preceded or followed a body movement by 5 seconds or more. Also eliminated was the EMG activity associated with spontaneous awakenings.

Results

The overall EMG rates for both deaf and normal groups were higher than had been the case in Part I, a result probably arising from the more sensitive amplifying devices employed in this second phase of the study. However, as may be seen in Table 3, the earlier finding of Part I was confirmed. Rates of finger EMG activity during REM periods in the deaf subjects certainly did not exceed those of the normal controls. If anything, the rates for normals were somewhat higher than those for the deaf individuals.

With regard to rates during non-REM sleep, it can be seen that the deaf subjects showed a slightly higher rate than did the normal controls. Nevertheless, there was a high degree of overlap between the two distributions, and the ranges for the two groups were roughly the same.

TABLE 3

Finger EMG Rate in Deaf versus Hearing Subjects: Part II

Phase of sleep cycle	Deaf		Hearing	
	M	Range	*M*	Range
Spindles + 15	1.16	.20–2.93	1.43	.33–4.07
Stages 2, 3, 4	.94	.42–1.80	.71	.17–1.73
15 Pre-REM	1.24	.33–2.09	.79	.17–1.60
REM Period	2.08	1.42–2.76	2.48	2.00–3.19
15 Post-REM	1.06	.27–1.98	.69	.36–2.79
Time Asleep	417 min.	369 min. 447 min.	413 min.	353 min. 456 min.
Subject *N*		6		6
Night *N*		6		6

Note.—Calculated as rate per minute.

Concerning the general pattern of EMG activity during the sleep cycle, it can be seen that for both deaf and hearing groups, the rate dropped after the first 15 minutes of sleep, remained low during non-REM sleep, then rose sharply at the onset of REM periods. This pattern of highest EMG activity during REM periods held for each of the 12 subjects ($p < .01$). Following REM periods the rate again dropped sharply.

The picture of EMG activity described here is similar to the pattern observed for body movement rates during sleep. Kamiya (personal communication), in 23 of his 25 subjects, noted a higher body movement frequency during REM periods, as opposed to the rate during other stages of sleep. Kamiya also observed that body movement rate, in the few minutes just preceding and in the few minutes just following a REM period, was even higher than in the REM period itself. Finger EMG activity, however, did not display this latter pattern.

It was also observed that the first REM period of the night showed consistently less finger EMG activity than did the succeeding REM periods. This was true for each of the 12 subjects ($p < .01$). The anomalous character of the first REM period complements data from other investigations. Shapiro and Goodenough (personal communication) have described first REM periods as having fewer eye movements, being shorter in duration, and yielding less vivid dream reports, in comparison with later REM periods.

No other "time-of-night" effect was apparent. The second REM period, for example, exhibited a mean EMG rate just as high as that of later REM periods. And the rate of EMG activity in non-REM sleep was nearly as high early in the night as it was towards morning—although there were individual departures from this norm.

Discussion

As may be seen in Table 1, the mean frequency of recall from REM periods by the seven deaf subjects used in the first phase of the study was 73%. In the case of normal hearing subjects, recall from REM periods runs in the order of 80% (Dement & Kleitman, 1957b) to 85% (Kamiya, 1961). The somewhat lower mean rates of 73% obtained for the deaf group in this study was attributable mainly to one of the subjects, JB, whose recall rate stood at 0%.

Probably the most reasonable interpretation of the recall data is that rates for the two groups were essentially similar. If anything, recall rates may have been somewhat lower for the deaf individuals. Certainly, the evidence offered no support for the statement of Mendelson, Siger, and Solomon (1960) that deaf subjects show superior recall of dreams. It should be noted, however, that the latter study did not employ the REM-EEG technique in combination with nocturnal awakenings, but instead used conventional postsleep recollections as the source of data.

Recall from non-REM sleep was 38%. Figures for non-REM recall in normal subjects vary considerably; Foulkes (1962) cited 74%; Kamiya (1961), 27%; Goodenough, Shapiro, Holden, and Steinschriber (1959), 35%; Rechtschaffen, Verdone, and Wheaton (1963), 23%. In any case, the recall rate of 38% by the deaf group fell within the normal range of variation.

The major point of interest at the outset of the study was whether the finger EMG activity of deaf subjects indicated the occurrence of non-REM mental activity. With each of the seven subjects, awakenings were conducted both: (*a*) immediately after the occurrence of finger EMG bursts, and (*b*) after finger EMG activity had been absent for several minutes. Analysis of the data showed that, contrary to expectations, non-REM recall was equally good under both these conditions. In other words, there was no relationship between finger EMG activity and the recall of non-REM material.

With regard to the patterning and amount of EMG activity during the various stages of sleep, it was quite clear that an increased frequency of EMG bursts associated with finger movements occurred during REM periods. This observation helped to account for the high rate of dream recall (91%) which Max obtained from his EMG awakenings. It seems probable that Max carried out a high proportion of his EMG awakenings during Stage 1 REM periods. It is also probable that a high proportion of his non-EMG awakenings took place outside of REM periods. Consequently, he would have obtained a positive association between the presence of finger EMG and dream recall. The present study, however, indicated that Max could have obtained the same result for hearing subjects, since normals show precisely the same cyclical patterning of EMG activity. Perhaps the most reasonable interpretation of the increased EMG activity observed during REM periods is that it represents part of the generalized arousal pattern characteristic of this stage of sleep.

An unexpected result of the current investigation was its failure to confirm Max's work in one important aspect. Max reported a high rate of EMG activity in the deaf as opposed to a complete lack of such activity in hearing persons. This result could readily be interpreted as supporting a motor theory of thinking—in the sense that there is a rather specific relationship between thinking processes and minimal motor activities in the linguistic apparatus. However, the present study indicated no difference between deaf and normal subjects with respect to amount of finger EMG activity in any stage of sleep. Accordingly, finger movements during sleep on the part of deaf persons cannot be taken to substantiate the motor theory of thinking.

In the author's view, better evidence in favor of a motor theory of thinking comes from the REM-EEG studies of dreaming conducted in recent years. Correspondences between the type of dream content recalled and the pattern of eye movements immediately prior to the moment of awakening have been noted by Dement and Kleitman (1957b), and Roffwarg et al. (1962). Berger, Olley and Oswald (1962) found an absence of REMs in subjects with longterm blindness. These same individuals also reported a lack of visual imagery. In late-blinded subjects—who retained their visual imagery—REMs were present.

Studies of eye movements during waking states have been conducted by Antrobus, Antrobus, and Singer (1964). Here, also, there was evidence for a significant association between reported mental activity and the corresponding eye movement patterns.

REFERENCES

ANTROBUS, J. S., ANTROBUS, JUDITH S., & SINGER, J. L. Eye movements during daydreaming, visual imagery, and thought suppression. *Journal of Abnormal and Social Psychology,* 1964, **69**, 244–252.

ASERINSKY, E., & KLEITMAN, N. Regularly occurring periods of eye motility, and concomitant phenomena during sleep. *Science,* 1953, **118**, 273–274.

BERGER, R. J., OLLEY, P., & OSWALD, I. The EEG, eye movements and dreams of the blind. *Quarterly Journal of Experimental Psychology,* 1962, **14**, 183–186.

DEMENT, W., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (a)

DEMENT, W., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (b)

FOULKES, W. D. Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology,* 1962, **65**, 14–25.

GOODENOUGH, D., SHAPIRO, A., HOLDEN, M., & STEINSCHRIBER, L. A comparison of "dreamers" and "nondreamers": Eye movements, electroencephalograms, and the recall of dreams. *Journal of Abnormal and Social Psychology,* 1959, **59**, 295–302.

JACOBSON, E. *Progressive relaxation.* Chicago: Univer. Chicago Press, 1929.

JACOBSON, E. Electrophysiology of mental activities. *American Journal of Psychology,* 1932, **44**, 677–694.

KAMIYA, J. Behavioral, subjective and physiological aspects of sleep. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey, 1961. Pp. 145–174.

MAX, L. W. An experimental study of the motor theory of consciousness: III. Action-current responses in deaf-mutes during sleep, sensory stimulation, and dreams. *Journal of Comparative Psychology,* 1935, **19**, 469–486.

MAX, L. W. An experimental study of the motor theory of consciousness: IV. Action-current responses in the deaf during awakening, kinesthetic imagery, and abstract thinking. *Journal of Comparative Psychology,* 1937, **23**, 301–304.

MENDELSON, J. H., SIGER, L., & SOLOMON, P. Psychiatric observations on congenital and acquired deafness: Symbolic and perceptual processes in dreams. *American Journal of Psychiatry,* 1960, **116**, 883–888.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford, 1953.

RECHTSCHAFFEN, A., VERDONE, P., & WHEATON, J. Reports of mental activity during sleep. *Canadian Psychiatric Association Journal,* 1963, **8**, 409–414.

ROFFWARG, H., DEMENT, W., MUZIO, J., & FISHER, C. Dream imagery: Relationship to rapid eye movements of sleep. *A.M.A. Archives of General Psychiatry,* 1962, **7**, 235–258.

WOLPERT, E. *An electromyographic study of dreaming.* (Doctoral dissertation, University of Chicago) Chicago: Documentation Center, 1959.

WOLPERT, E. Studies in the psychophysiology of dreams: II. An electromyographic study of dreaming. *A.M.A. Archives of General Psychiatry,* 1960, **2**, 231–241.

(Received September 12, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 350–359

PURSUIT-ROTOR PERFORMANCE IN THE BRAIN DAMAGED[1]

MARY M. HUSE[2] AND OSCAR A. PARSONS

Duke University Medical Center

University of Oklahoma Medical Center

Highly significant differences in level of performance and rate of improvement on the pursuit-rotor were found between Brain Damaged (BD, N = 40) and Control (N = 40) patients under massed and spaced practice conditions, but no differential deficit due to conditions was present. No differences were found between mildly impaired BD (independently rated for cerebral impairment) and Control Ss on measures of reactive inhibition (Ir) and conditioned inhibition (sIr), although these groups differed in performance level and improvement rate. Massed practice curves for severely impaired BD Ss were markedly different but not attributable to disturbances in Ir and sIr. Impairment of new learning and sensory-motor dysfunction rather than disturbances in inhibition seem salient in pursuit-rotor performance deficit of the brain damaged.

There is rather widespread agreement that perceptual-motor and learning functions are affected adversely by brain injury. However, there are surprisingly few studies, using brain-damaged subjects, in which task and subject variables have been explored in relation to deficit, particularly in the area of motor skills. The pursuit-rotor task would appear to have many advantages for the investigation of deficit performance in the brain damaged: it is less sensitive to socioeducational factors than many tasks; the parameters governing performance have been extensively investigated; some of the theoretical constructs which have been used to predict performance have been related to central nervous system functions; and performance on such a task might provide relevant information for diagnostic and rehabilitative efforts.

The first aim of the present study was to compare the levels of performance and rates of improvement of brain-damaged (BD) subjects and non-brain-damaged (NBD) subjects on the pursuit-rotor under conditions of massed and spaced practice. Changes in somatosensory functions (Semmes, Weinstein, Ghent, & Teuber, 1960), perception (Teuber & Bender, 1949), memory (Graham & Kendall, 1946) and attention (Reitan & Kløve, 1959) are recognized as important factors in performance deficit. As these functions enter into performance on a pursuit-rotor it would be expected that the level of performance on this task would be lower for BD subjects than for NBD subjects. In addition, because injury to the brain has been shown to affect the acquisition of new habits, (Gantt, 1938; Hall & Crookes, 1951; Reese, Doss, & Gantt, 1953) the rate of improvement would be expected to be lower for the BD group as compared with the NBD group. Consideration of Reitan's (1959) statement, "Tests most sensitive to any organic impairment of brain functions are ones requiring alertness, concentrated and sustained attention to the task . . . [p. 18]" would suggest that a relatively greater decrement in performance would be found under massed practice compared with spaced practice conditions in the BD group.

The second aim of the present study was to examine the performances of the BD subjects under the above conditions with respect to the extent or severity of brain damage. Strong evidence has been provided (Chapman, Thetford, Berlin, Guthrie, & Wolff, 1958) that degree of performance deficit is proportional to the amount of brain tissue lost or destroyed.

---

[1] This report is based on a dissertation by the first author, submitted in partial fulfillment of the requirements for PhD degree in the graduate school of Duke University. The study was conducted within the context of a larger program of research under the direction of the second author and supported, in part, by United States Public Health Service Grant B-2507.

[2] The authors are indebted to C. A. Boneau, L. D. Cohen, G. Kimble, and T. Peele of Duke University for technical assistance in various phases of this project, and to Betty Chandler and Helen Klein of the University of Oklahoma for their editorial assistance. Special thanks are due Paul Costiloe of the University of Oklahoma Medical Research Computer Center for help in processing these data.

The final aim was an exploration of the extent to which inhibitory variables may account for differences in motor learning in the brain damaged. Under massed practice conditions a decline in performance with practice has been attributed to the accumulation of reactive inhibition (Ir), and an improvement in performance after rest, that is, reminiscence, has been attributed to the dissipation of Ir (Hull, 1943; Kimble, 1949a; Kimble, 1949b; Kimble & Horenstein, 1948). Learned or conditioned inhibition (sIr) is postulated to occur after a critical point in Ir buildup has occurred and results in a decreasing slope of the learning curve and no decline in performance (Hull, 1943; Kimble, 1949a). Concepts of reactive inhibition (Ir) and conditioned inhibition (sIr) have predicted such empirical phenomena as the differential effects of massed practice as compared with spaced practice, decline, and reminiscence (Hull, 1943; Kimble, 1949a; Bilodeau & Bilodeau, 1961). In view of the fact that many investigators have related Ir and sIr to some type of central nervous system (CNS) construct (Bilodeau & Bilodeau, 1961) it would appear that disturbances in CNS inhibitory processes as a function of brain injury might be expected to affect scores for decline, reminiscence, and sIr.

In summary, the aims of the present experiment were: (*a*) to compare the performances of brain-damaged and non-brain-damaged subjects on the pursuit-rotor under conditions of massed and spaced practice, as regards level and rate of improvement; (*b*) to relate severity of damage to performance deficit; and (*c*) to investigate the extent to which Hull's concepts of reactive (Ir) and conditioned (sIr) inhibition may account for differences in pursuit-rotor performance in brain damage.

METHOD

Four groups of hospitalized subjects ($N = 20$ per group) were given a pursuit-rotor task: a control spaced practice group; a control massed practice group; a brain-injured spaced practice group; and a brain-injured massed practice group.

Subjects

BD and NBD male patients were recruited from the Duke Hospital public wards, the Duke Hospital Brain Tumor Clinic, and the Durham, North Carolina Veterans Administration Hospital. The groups were equated, insofar as possible, on the following variables (see Table 1): age, race, education, WAIS Vocabulary IQ, anxiety, motivation as rated by the experimenter in the task situation, handedness (90% right handed), the proportion taking cortical depressant medication, and the presence of psychiatric difficulties. The only significant difference on the control measures was found between the organic and control spaced practice groups in the proportion of patients taking cortical depressants.

NBD groups were comprised of patients who had no neurological symptoms and histories free of a variety of diseases which might affect cerebral functioning. Psychiatric, psychophysiological, surgical, and medical patients constituted these groups. Criteria for inclusion in these groups were:

> Medical histories free of the following: head injury, prolonged unconsciousness, seizures, cerebral

TABLE 1

STANDINGS OF THE SUBJECT GROUPS ON THE CONTROL MEASURES

Group	Age	Education	IQ	Anxiety	Motivation	Psychiatric problems	Proportion on cortical depressants
Control spaced						7	.200
$\bar{X}$	34.25	9.65	92.90	18.05	6.10		
SD	7.43	2.28	16.41	8.23	.79		
Control massed						7	.300
$\bar{X}$	35.40	10.33	99.00	17.50	5.85		
SD	7.60	3.73	21.00	9.15	.92		
Organic spaced						7	.750
$\bar{X}$	35.20	8.95	88.70	21.50	6.10		
SD	7.01	3.71	16.45	9.32	1.02		
Organic massed						6	.550
$\bar{X}$	35.00	9.58	92.40	16.70	6.15		
SD	6.92	2.67	11.76	8.93	1.00		

TABLE 2

LOCI OF LESIONS IN THE TWO BRAIN-INJURED GROUPS

Spaced practice group	Massed practice group
1. Resection from anterior pole, right temporal lobe	Right temporal EEG focus
2. Right occipital EEG focus	Suspected brain tumor, right internal capsule
3. Left mid-temporal EEG focus	Left mid-temporal EEG focus
4. Right temporal EEG focus	Ruptured aneurysm, right middle cerebral artery, and hematoma right temporal lobe
5. Intracerebral angioma involving entire right hemisphere	Three aneurysms of right middle cerebral artery with sub-arachnoid bleeding
6. Left temporal EEG focus	Head injury with residual large defect left temporal area
7. Left temporal EEG focus	Left temporal EEG focus
8. Right occipito-parietal brain	Right parietal intracerebral cyst
9. Left motor area CVA	Ruptured aneurysm, left motor area
10. Gunshot wound, left frontal area	Gunshot wound, left frontal area
11. Generalized brain damage	Trauma to right and left fronto-parietal area of skull, with generalized EEG abnormalities
12. Generalized abnormal EEG	Generalized abnormal EEG
13. Bilateral temporal EEG abnormalities	Bilateral temporal subdural hygromas
14. Brain abscess, left parietal area	Left parietal astrocytoma
15. Left motor area CVA	Left temporal EEG abnormality
16. Right motor area CVA	Gunshot wound, right motor area
17. Head trauma, right motor area	Right motor area CVA
18. Left temporal lobectomy	Left parieto-occipital brain tumor
19. Right front-temporal brain tumor	Possible right frontal brain tumor
20. Centrencephalic EEG focus	Centrencephalic EEG focus

> vascular accident, blood dyscrasias, pernicious anemia, long-standing and uncontrolled diabetes, long-standing and uncontrolled hypertension, chronic and severe cardiac difficulties, chronic and severe lung difficulties, chronic and severe renal difficulties, severe endocrine disturbances, multiple sclerosis, and parkinsonism.

BD subjects were diagnosed by neurologists and neurosurgeons and were included in the present study if they met one or more of the following criteria:

> The presence of a brain tumor or a previous operation for this condition; cerebral aneurysm; cerebral vascular accident, with subsequent positive neurological findings (including not only the results of clinical neurological examination, but also the results of EEG tests, air studies, and spinal taps); head injury with subsequent positive neurological findings; suspected cerebral lesion, supported by positive neurological findings; cortical atrophy; and seizures accompanied by abnormal EEG findings.

The BD spaced and massed practice groups were matched on a subject by subject basis for right, left, or generalized lesion (see Table 2). The groups also were divided into severely and mildly impaired BD on the basis of a five-point scale of severity of impairment of cerebral functions (independent of the experimental results) with 85% agreement among three judges (the authors and a board neurologist) as to categorization.[3] Of the 18 severely impaired subjects, 9 fell into the massed practice group and 9 into the spaced practice group. Eleven of the mildly impaired subjects fell into each of the two groups. These groups did not differ significantly on any of the control variables.

Apparatus

The apparatus was a modified Koerth pursuit-rotor which ran in a counter-clockwise direction at the rate of 64 revolutions per minute. The turntable was $11\frac{3}{4}$ inches in diameter and was painted black. A round brass target, the size of a quarter, was placed flush with the surface $1\frac{7}{8}$ inches from the outside edge of the turntable. The task was to keep the point of a flexible brass stylus in contact with the brass target as much as possible. Contact with the target closed an electric circuit and activated one of two standard electric timers in the circuit. These timers were activated alternately, and the experimenter recorded the time on target every 30 seconds. This arrangement permitted the subjects to work without interruption during the massed work periods.

Procedure

The work-rest schedules for the spaced and massed practice groups allowed for a total of 15 working minutes for each group. These schedules had been found to be optimal in preliminary experimentation with normal subjects. The spaced practice group had 30 second work periods and 1-minute rest periods alternately for 30 trials. The massed practice group had four work periods which were scheduled as follows: 1 minute 30 seconds work; 6 minute rest; 4 minutes 30 seconds work; 6 minutes rest; 4 minutes 30 seconds work; 6 minutes rest; and 4 minutes 30 seconds work. Work during the massed trials was uninterrupted. The initial work period of 1 minute 30 seconds, the short work period, made possible the examination of differences in the performance curves of the subjects during and after short and long work periods.

---

[3] For a detailed description of criteria and neurological ratings see Huse (1959).

When the subject entered the testing room the experimenter chatted with him until he appeared at ease. Standard instructions about the task then were given.

Throughout the testing every effort was made to make certain that the subjects understood the task and that they were motivated to do their best. Directions were repeated as often as necessary. Scores were reported to the subjects and they were encouraged about them. Absolute standardization of procedures was sacrificed in view of the deemed importance of maintaining motivation.

During the rest periods the Vocabulary subtest from the Wechsler Adult Intelligence Scale (WAIS) and the Taylor Manifest Anxiety scale (*MA*) were administered to all groups.

Derivation of Scores for Ir and sIr Variables

Decline scores during the short work period were calculated by finding the difference between a subject's score on Trial 3 and his score on Trial 1. Decline scores during the long work periods were calculated by subtracting the subject's extrapolated last-trial score from his extrapolated first-trial score on the basis of the straight line functions described below. A minus score would mean that, rather than decline, improvement in performance had occurred.

Reminiscence after the short work period was calculated by fitting a straight line function by the method of least squares (Lewis, 1948) to the last eight trials of the first long work period. This curve was then extrapolated backward to establish the first postrest point from which was subtracted the mean of the subject's first three trials. This constituted the reminiscence score for the short work period. Reminiscence scores after the long work periods were derived similarly by subtracting the extrapolated last prerest point from the extrapolated first postrest point of the next massed trial.

The sIr scores were calculated by fitting Hull's exponential curve to the empirical spaced practice performance curve; then each subject's extrapolated first postrest trial score of the massed practice condition was subtracted from the fitted score of the matched spaced practice group on the same trial (Lewis, 1948, pp. 61–64).

RESULTS

The spaced and massed practice curves for the NBD and BD groups are presented in Figure 1. The data on which these curves are based were analyzed by a mixed design analysis of variance (Lindquist, 1956; see Table 3). It can be seen that the BD performed more poorly than the NBD under both spaced and massed practice conditions ($p < .01$).

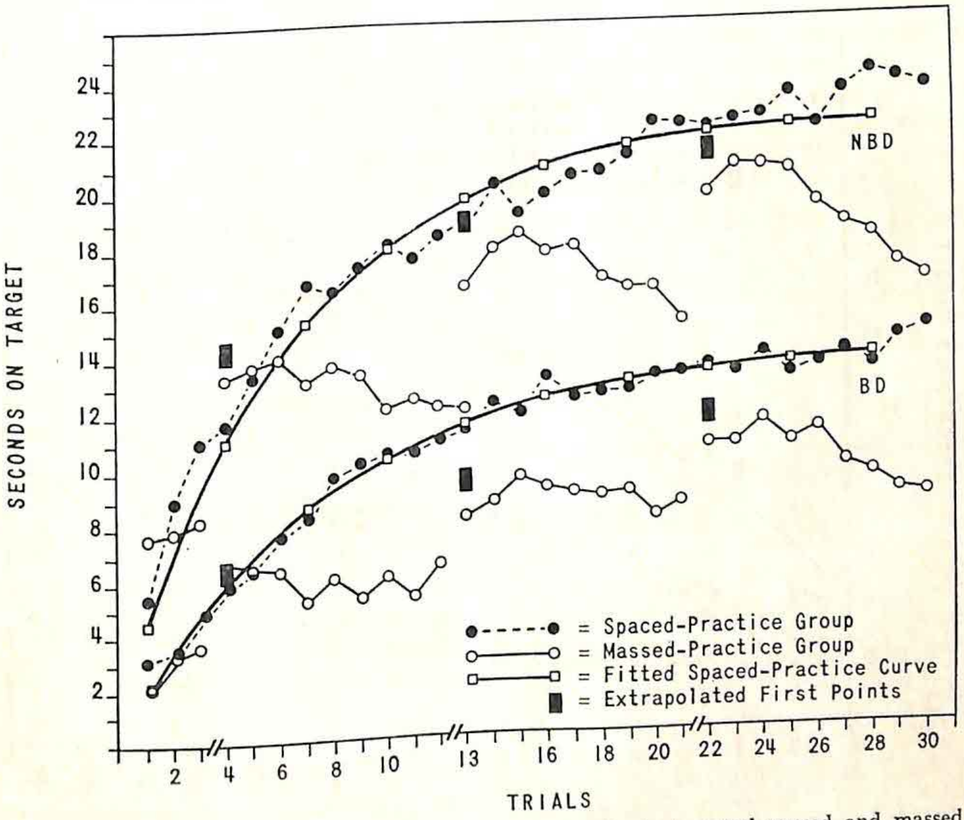


FIG. 1. Learning curves for the non-brain-damaged and brain-damaged spaced and massed practice groups.

TABLE 3

Analysis of Variance of Pursuit-Rotor Performance of BD and NBD Groups under Spaced and Massed Practice Conditions, 30 Trials

Source	*df*	*MS*	*F*	*p*
Between subjects	79			
BD versus NBD (A)	1	36,754.02	51.20	<.01
Spaced versus Massed (B)	1	5,900.06	8.22	<.01
Group × Treatment (A × B)	1	.49	.00	ns
Error	76	71,779.64		
Within subjects	2,320			
Trials (C)	29	91,905.93	138.58	<.01
Group × Trials (A × C)	29	4,258.24	6.42	<.01
Treatment × Trials (B × C)	29	8,373.34	12.62	<.01
Group × Treatment × Trials (A × B × C)	29	785.38	1.18	ns
Error	2,204	663.20		
Total	2,399			

BD also had a lower rate of improvement than NBD, as determined by the Groups × Trials (A × C) interaction. However, the differences in level of performance between BD and NBD under spaced and massed practice conditions were about the same (nonsignificant A × B interaction) and did not differ over trials (A × B × C). Apparently massed practice did not have a more deleterious effect than spaced practice for the BD as compared with NBD.

A second aim of the present study was to compare the levels of performance and rates of improvement for the mildly and severely impaired subjects. Performance curves are presented in Figure 2. As expected, the severely impaired subjects performed more poorly and had a lower rate of improvement than the

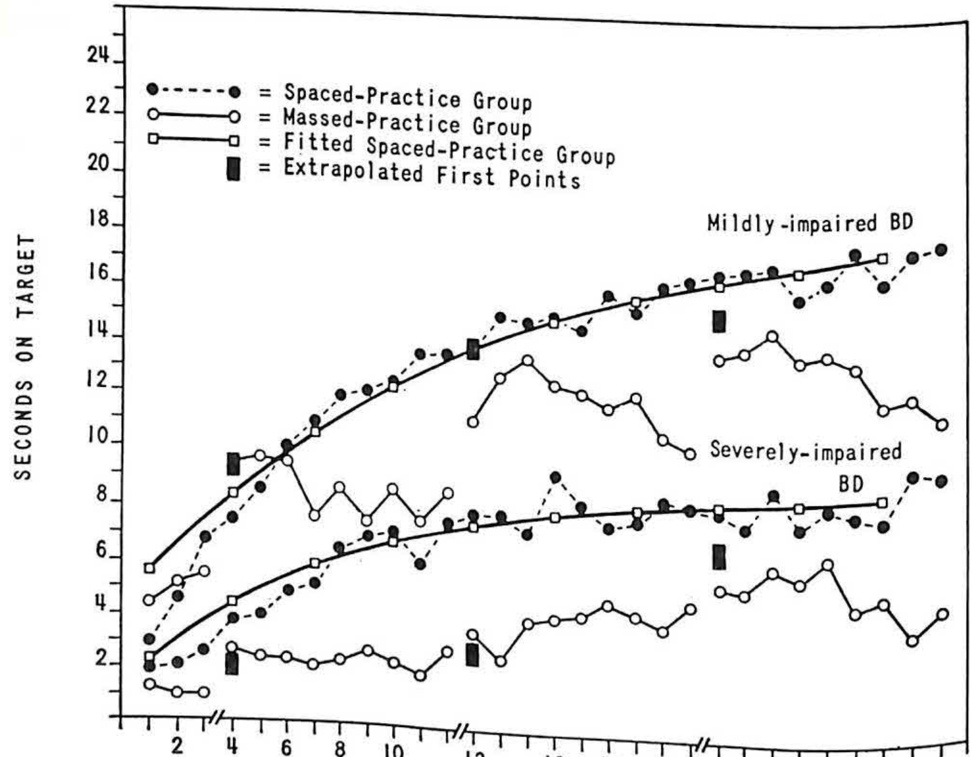


Fig. 2. Learning curves for the mildly impaired and severely impaired spaced and massed practice groups.

mildly impaired subjects. Tests for differences on the performance scores were made by comparing the mildly impaired and severely impaired groups with each other and with the control group. The Welch-type comparison of means was used (Welch, 1951) because of its appropriateness for groups of unequal size and variance. Significant differences are present for both spaced and massed practice conditions (see Table 4).

Because the appropriateness of *t* tests is questionable when the sample sizes and variances differ, the *U* test (Siegel, 1956), was used to test for differences between the groups (see Table 5). Under both the spaced and massed practice conditions the performance of the NBD group was significantly better than that of the mildly impaired BD group ($p < .02$) and both these groups were better than the severely impaired BD group ($p < .001$). Again it may be noted (Table 4) that differences in level of performance among the three groups were about the same for both spaced and massed practice conditions. Finally, *U* test comparisons of the rate of improvement (last 2 minutes minus first 2-minutes performance) for the severely and mildly impaired groups and the NBD group were made. The spaced practice improvement means (in seconds) for these groups were, respectively, 6.5, 12.3, 14.6; the corresponding massed practice means were 3.2, 6.6, 8.6. The severely impaired BDs were significantly lower on rate of improvement than either of the other groups but the mildly impaired were not significantly different from the NBD (see Table 6).

Since the last 2 minutes minus first 2 minutes is a relatively crude measure, the rate of improvement in the mildly impaired was investigated more thoroughly. NBD subjects ($N = 22$) were matched to the 22 mildly impaired BD on the basis of their first trial performance. These groups were then compared over 30 trials by analysis of variance. Significant differences were found in both level of performance ($F = 4.76$, $df = 1/40$, $p < .05$) and rate of improvement ($F = 2.02$, $df = 29/1{,}160$, $p < .01$).

It is concluded that severely impaired BD, mildly impaired BD and NBD differ from each other in level of performance and rate of improvement. The more severe the brain damage the lower the level of performance and the lower the rate of improvement.

Tests for differences on aspects of performance reflecting inhibitory variables (Aim 3) were made by comparing the mildly and severely impaired BD with each other and with the NBD group. The Welch-type comparison of means and the *U* test were used to test for differences between groups (see Tables 7 and 8). In general, the scores on the various measures for the mildly impaired BD fell between those of the severely impaired BD and the NBD but did not differ from the latter. For the short work period, the severely impaired BD had significantly higher decline scores, lower reminiscence scores and higher

TABLE 4

WELCH-TYPE COMPARISON OF MEANS OF SEVERELY AND MILDLY IMPAIRED BD AND NBD ON PERFORMANCE, 30 TRIALS

Group	Spaced practice		Massed practice	
	M	Difference	*M*	Difference
Severely Impaired BD	7.13		3.80	
		6.60		7.19
Mildly Impaired BD	13.73		10.99	
		4.90		4.48
NBD	18.63		15.47	
V^2	21.06*		35.69*	

* $p < .01$.

TABLE 5

U TEST COMPARISONS OF SEVERELY AND MILDLY IMPAIRED BD AND NBD GROUPS ON PERFORMANCE, 30 TRIALS

Group	Spaced practice		Massed practice	
	U	*p*	*U*	*p*
Severely impaired versus mildly impaired BD ($N = 9, N = 11$)	10	$<.001$	3	$<.001$
Severely impaired BD versus NBD ($N = 9, N = 20$)	4	$<.001$	3	$<.001$
Mildly impaired BD versus NBD ($N = 11, N = 20$)	44.5	$<.02$	45	$<.02$

TABLE 6

U TEST COMPARISONS OF SEVERELY AND MILDLY IMPAIRED BD GROUPS AND NBD ON $\bar{X}$ LAST 2 MINUTES MINUS $\bar{X}$ FIRST 2 MINUTES OF PERFORMANCE

Group	Spaced practice		Massed practice	
	U	p	U	p
Severely impaired versus mildly impaired ($N = 9, N = 11$)	21	$<.05$	24	$= .05$
Severely impaired versus NBD ($N = 9, N = 20$)	12.5	$<.01$	15	$<.01$
Mildly impaired versus NBD ($N = 11, N = 20$)	82.0	$>.05$	81	$>.05$

TABLE 7

WELCH-TYPE COMPARISON OF MEANS OF SEVERELY AND MILDLY IMPAIRED BD AND NBD SUBJECTS ON DERIVED MEASURES

Group	Measure									
	Decline	Reminiscence	sIr	Decline			Reminiscence		sIr	
	SWP	SWP	SWP	LWP 1	LWP 2	LWP 3	LWP 1	LWP 2	LWP 1	LWP 2
Severely impaired BD	.50	1.16	2.07	− .35	−1.48	2.28	.61	2.04	4.31	1.82
Mildly Impaired BD	−1.05	4.36	− .98	1.31	2.96	3.54	5.99	2.51	−.05	1.19
NBD	− .50	6.58	−3.21	2.11	3.29	4.67	6.81	6.19	.78	.08
V^2	1.37	11.58**	8.88**	2.52	7.13**	1.88	10.24**	5.71*	2.92	.47

* $p < .05$.
** $p < .01$.

sIr scores than the mildly impaired BD and the NBD groups.

For the long work periods, the severely impaired BD group had significantly lower decline scores than the other groups on the second long work period only, but significantly lower reminiscence scores than the mildly impaired BD and the NBD on three of the four possible comparisons (Table 8). No differences were found among groups on sIr measures for the long work periods, although the severely impaired BD had consistently higher scores than the other two groups.

DISCUSSION

The results give ample evidence that pursuit-rotor performance was impaired both in terms of level of performance and rate of improvement in a heterogeneous group of BD subjects. The task variable of spaced versus

TABLE 8

U TEST COMPARISONS OF SEVERELY AND MILDLY IMPAIRED BD AND NBD GROUPS ON THE DERIVED MEASURES

Group	Measure									
	Decline	Reminiscence	sIr	Decline			Reminiscence		sIr score	
	SWP	SWP	SWP	LWP 1	LWP 2	LWP 3	LWP 1	LWP 2	LWP 1	LWP 2
Severely versus Mildly Impaired BD ($N = 9, N = 11$)	15*	22*	21*	35	16*	38	11*	29	29	43
Severely Impaired BD versus NBD ($N = 9, N = 20$)	50*	20**	34*	52	27**	62	21**	33*	54	72
Mildly Impaired BD versus NBD ($N = 11, N = 20$)	52*	75	82.5	104	98	100	87	83	112	86

* $p \leq .05$.
** $p < .01$.

massed practice, surprisingly, gave rise to no differential deficit in the BD as compared with NBD. The subject variable of independent ratings of impairment of cerebral functions led to evidence that severely impaired BD differ markedly both in level of performance and rate of improvement from the other two groups, under both massed and spaced practice conditions. The mildly impaired BD were significantly lower in level of performance and rate of improvement than the NBD.

Level of Performance and Rate of Improvement

The sensitivity of the pursuit-rotor to disturbance in brain functions is emphasized by the fact that the severely impaired BD, mildly impaired BD, and NBD were significantly different from each other, in the expected direction, on the first 30 seconds trial (see Table 9). This initial difference in level poses a problem of interpretation of the findings for rate of improvement. Rather than impaired motor learning, the lower rate of improvement may reflect a changed performance curve due to impaired sensory motor functioning. Analysis of performance in relation to laterality of lesion should be helpful here. With right-handed subjects it would be expected that greater performance deficit would be found in the left-hemisphere lesion cases than in the right-hemisphere cases. If these groups did not differ in rate of improvement it could be inferred that motor learning was impaired despite differences due to sensory-motor dysfunction.

To test these ideas four groups of right-handed BD subjects were composed (who were not grossly different on any of the control variables): lesion predominantly right-sided, spaced practice ($N = 8$), massed practice ($N = 8$); lesion predominantly left-sided, spaced practice ($N = 8$), massed practice ($N = 8$). Analysis of variance indicated that differences between right- and left-hemisphere cases on spaced and massed practice $\overline{X}$ level of performance, while in the expected direction (right-sided spaced $\overline{X} = 13.2$ seconds, massed $= 9.4$ seconds; left-sided spaced $X = 10.5$ seconds, massed $\overline{X} = 6.2$ seconds), were not significant (see Table 10). More importantly, however, there was no difference between left- and right-sided lesions cases in rate of improvement (A × C and A × B × C interactions, see Table 10). Since the Ns were small and included the severely impaired it was decided, as a further check, to look at the group who should have the least sensory-motor dysfunction, the mildly impaired, right-hemisphere cases. Eleven NBD subjects were matched on first trial performance with 6 mildly impaired spaced practice and 5 mildly impaired massed practice subjects. Rates of improvement in terms of last 2 minutes minus first 2 minutes were then compared. In both groups BD had significantly lower rates of improvement than the NBD (spaced $U = 4$, $p < .01$; massed $U = 4$, $p < .05$).

In considering these results, there is little doubt that impairment of motor learning contributes substantially to the performance deficit of the BD. The role of sensory-motor dysfunction remains to be explored more thoroughly; however, the evidence given above suggests that quantification of techniques designed to systematically measure sensory and motor functions may prove valuable. Such an approach would be consistent with that of Reitan and Kløve (1959) who have demonstrated the sensitivity of finger tapping and tactual performance tests, when carefully quantified, to lesion location and lateralization, and our own successful identification of brain lesions on the basis of flicker perimetry (Parsons & Huse, 1958).

TABLE 9

U TESTS OF THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN GROUPS IN PERFORMANCE ON FIRST TRIAL (30 SECONDS)

Group	Spaced practice		Massed practice	
	U	p	U	p
Severely Impaired versus Mildly Impaired BD ($N = 9$, $N = 11$)	43.5	ns	20.0	$<.05$
Severely Impaired BD versus NBD ($N = 9$, $N = 20$)	30.5	$<.02$	16.0	$<.002$
Mildly Impaired BD versus NBD ($N = 11$, $N = 20$)	58.5	$<.05$	60.0	$<.10$

TABLE 10

Analysis of Variance of Pursuit-Rotor Performance for Right- and Left-Sided Lesion Groups

Source	df	MS	F	p
Between subjects	31			
Right versus Left (A)	1	240,666.	2.64	ns
Spaced versus Massed (B)	1	352,283.	3.86	ns
Groups × Treatments (A × B)	1	11.	.00	ns
Error	28	91,324.85		
Within subjects	928			
Trials (C)	29	24,169.86	33.22	<.01
Groups × Trials (A × C)	29	447.59	.62	ns
Treatments × Trials (B × C)	29	3,084.90	4.24	<.01
Groups × Treatments × Trials (A × B × C)	29	407.79	.56	ns
Error	812	727.47		
Total	959			

Spaced versus Massed Practice and Inhibition

One of the most puzzling findings in this study is the lack of differential deficit in BD and NBD under massed compared with spaced practice. On almost any basis of reasoning, the BD should be maximally different from NBD under massed practice. As noted earlier the pursuit-rotor massed practice condition would seem to fit Reitan and Kløve's (1959) description of test situations which are most sensitive to brain dysfunction. Chapman and Wolff (1959) have described changes in the goal-oriented behavior of BD subjects, under conditions where sustained effort is required, which would lead to the deficit we predicted. At the theoretical level a number of investigators, particularly Eysenck, have related Ir to some type of CNS function (Duncan, 1956; Eysenck, 1955, 1956). Greater build up of Ir, more rapid development of sIr and consequent greater deficit in performance in the BD would be expected under massed practice as compared with spaced practice. It is, then, of particular interest to ask why this effect was not obtained, particularly in our mildly impaired BD groups. (The severely impaired were functioning at such a generally low level that the effects of variables which might depress performance may have been obscured by a "floor" phenomenon.)

One plausible explanation is in terms of motivation. It will be recalled that Hull (1943) maintained that Ir and sIr subtract from the motivational variable D in the sHr × D multiplicative equation. Now as noted in the method section, it was deemed important to keep the patients well motivated in this experiment, especially the BD. Special steps were taken to ensure this, for example, reporting of scores, continual supportive and encouraging remarks, etc. As seen in Table 1 the average for all groups is around six on a nine-point scale of motivation, indicating "above average cooperation and effort to do well." From this point of view our motivated mildly impaired BD subjects were able to tolerate as much Ir as NBD and hence were similar in sIr development. In essence, then, we propose that the performance deficit on the pursuit-rotor in the BD lies on the sHr part of the equation, and in the sensory-motor abilities which underlie learning, rather than on the D side.

While this explanation may hold for the mildly impaired groups, which were not significantly different from the NBD on the various measures of inhibition, what about the severely impaired BD who differed on many? It seems unlikely that these differences are attributable to processes which Ir and sIr are commonly understood to represent. Impairment of cerebral function in these cases appeared to be so great as to preclude these subjects from achieving even a minimal level of performance until about halfway through the test. Note also the very atypical first two massed practice curves (see Figure 2) of the severely impaired. The first curve is flat, the second gently rising and it is not until the third massed practice work period that a curve resembling those of the mildly impaired and NBD appears. This finding also argues against another possible interpretation of these differences of the severely impaired, that is, as the consequence of the early development of sIr on the first several trials. While the latter explanation could account for the differences under the first two massed practice conditions it would be difficult to explain the third massed practice curve in which the more typical warm-up reminiscence and decline, usually attributed to Ir, appear. Apparently a certain degree of task competency must be achieved before the operation of variables such as Ir and sIr are demonstrable. It would be of interest in this respect to carry the performance cycle for the severely brain damaged for several more massed practice trials. Would the more typical pursuit-rotor performance curves be found? Would reminiscence and decline increase and then decrease again?

In conclusion, it seems reasonable to infer that whatever the neurophysiological processes are which underlie the phenomena captured by the constructs Ir and sIr, they are minimally, if at all, affected by brain damage. This is of interest in light of the recent conclusion by Ammons (1960) reaffirming the importance of "central phenomena" in both types of inhibition. What are these "central phenomena"? Rather than equating Ir with some type of neurophysiological build-up in specific cortical structures as a result of excitation of those structures, as Eysenck (1955) seems to do, it may well be that more molar neurophysiological variables, for example, level of activation, are involved. It is probably no accident that Ir and sIr are demonstrated best in performance tasks calling for repetitive activity, where habituation can easily occur and lowered levels of arousal ensue.

REFERENCES

AMMONS, C. H. Temporary and permanent inhibitory effects associated with acquisition of a simple perceptual-motor skill. *Journal of General Psychology,* 1960, **62**, 223–245.

BILODEAU, E. A., & BILODEAU, INA McD. Motor skills learning. *Annual Review of Psychology,* 1961, **12**, 243–280.

CHAPMAN, L. F., THETFORD, W. N., BERLIN, L., GUTHRIE, T. C., & WOLFF, H. G. Highest integrative functions in man during stress. In H. Solomon, S. Cobb, and W. Penfield (Eds.), *The brain and human behavior.* Baltimore: Williams and Wilkins, 1958.

CHAPMAN, L. F., & WOLFF, H. G. The cerebral hemispheres and the highest integrative functions of man. *A. M. A. Archives of Neurology,* 1959, **1**, 357–424.

DUNCAN, C. P. On the similarity between reactive inhibition and neural satiation. *American Journal of Psychology,* 1956, **69**, 227–235.

EYSENCK, H. J. Cortical inhibition, figural aftereffect and theory of personality. *Journal of Abnormal and Social Psychology,* 1955, **51**, 94–106.

EYSENCK, H. J. Reminiscence, drive and personality theory. *Journal of Abnormal and Social Psychology,* 1956, **53**, 328–333.

GANTT, W. H. A method of testing cortical function and sensitivity of the skin. *A. M. A. Archives of Neurology and Psychiatry,* 1938, **40**, 79–85.

GRAHAM, FRANCES K., & KENDALL, B. S. Performance of brain-damaged cases on a memory-for-designs test. *Journal of Abnormal and Social Psychology,* 1946, **41**, 303–314.

HALL, K. R. L., & CROOKES, T. G. Studies in learning impairment: I. Schizophrenia and organic patients. *Journal of Mental Science,* 1951, **97**, 729–737.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

HUSE, MARY M. Pursuit-rotor performance in the brain-damaged: The role of reactive and conditioned inhibition in cortical dysfunction. Unpublished doctoral dissertation, Duke University, 1959.

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *Journal of Experimental Psychology,* 1949, **39**, 15–23. (a)

KIMBLE, G. A. A further analysis of the variables in cyclical motor learning. *Journal of Experimental Psychology,* 1949, **39**, 332–337. (b)

KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *Journal of Experimental Psychology,* 1948, **38**, 239–244.

LEWIS, D. *Quantitative methods in psychology.* Iowa City: Gordon Bookshop, 1948.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

PARSONS, O. A., & HUSE, MARY M. Impairment of flicker discrimination in brain-damaged patients. *Neurology,* 1958, **8**, 750–755.

REESE, W. G., DOSS, H., & GANTT, W. H. Autonomic responses in the differential diagnosis of organic and functional psychoses. *A. M. A. Archives of Neurology and Psychiatry,* 1953, **70**, 778–793.

REITAN, R., & KLØVE, H. The effects of brain lesions on adaptive abilities in human beings. Indianapolis: Indiana University Medical Center, 1959. (Mimeo)

SEMMES, JOSEPHINE, WEINSTEIN, S., GHENT, LILA, & TEUBER, H. L. *Somatosensory changes after penetrating brain wounds in man.* Cambridge: Harvard University Press, 1960.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TEUBER, H. L., & BENDER, M. Alterations in pattern vision following trauma in the occipital lobes in man. *Journal of General Psychology,* 1949, **40**, 38–57.

WELCH, B. L. On the comparison of several mean values: An alternative approach. *Biometrika,* 1951, 38, 330–336.

(Received September 18, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 360–370

ATTRACTION AND AVERSION TO POWER:

A PSYCHOANALYTIC CONFLICT THEORY OF HOMOSEXUALITY IN MALE PARANOIDS[1]

HOWARD M. WOLOWITZ
University of Michigan

Consideration of psychoanalytic views of paranoid symptomatology led to the assumption of an approach-avoidance conflict in male paranoids regarding power in other men. Approach involved wishes to be dependent on powerful men and to appropriate their power through magical sexual-aggressive means. Avoidance involved anticipated fears of retaliatory injury and destruction by more powerfully perceived males. A deduction that increments in the power of males, but not females, would result in paranoids' decreased resultant attraction was confirmed. The traditional, undifferentiated like-sex attraction hypothesis was not supported.

PSYCHOANALYTIC THEORIES OF PARANOIA

Traditional psychoanalytic theory of the dynamics of paranoid symptomatology attributed a basic role to a vigorous defense against an impelling homosexual attraction (Freud, 1950c).

Over the following years a sizeable number of confirmatory clinical cases were reported (e.g., Brill, 1934; Devine, 1916, 1921; Ferenczi, 1955, 1950b, 1950c; Freud, 1950a, 1950d, 1950b; Frink, 1926; Shockley, 1914, etc.). Less well-known are an impressive number of experimental confirmations (Aronson, 1950; Daston, 1952; Gardner, 1931; Klaf & Davis, 1960; Moore & Selzer, 1963; Musiker, 1952; Norman, 1948; Page & Warkentin, 1938; Zamansky, 1953).

In spite of the clinical and experimental attention given to the homosexual conflict hypothesis, the quality of the component attractive and aversive forces never received the systematic study they deserved although a number of theoretical suggestions have been offered. Early contributors, committed to libido theory, stressed submissive, anal-receptive wishes for, and fear of anal-penetrative assault by, the father or father substitute (Ferenczi, 1955, 1950b, 1950c; Stärke, 1920; Von Ophuijsen, 1920).

These observations and theoretical constructions raised the following questions: "Why does the male paranoid develop such intense submissive-receptive wishes for the father?" and secondly, "Why are they repudiated so fiercely?"

Freud (1950b) suggested the following explanation of the male child's repudiation of feminine wishes for the father:

> The feminine attitude to the father became repressed as soon as the boy realized that his rivalry with the woman for his father's love implies the loss of his own male genital, that is to say, implies castration. Repudiation of the feminine attitude is therefore a result of the struggle to avoid castration; it regularly finds its most emphatic expression in the fantasy of castrating the father and turning him into a woman [p. 455].

Although intended as a general explanation for the repudiation of receptive wishes for the father, it follows that the male paranoid's fear of his own receptive-submissive sexual wishes for other males may be based on similar, but more intensive, anticipations. Freud (1950b) assumed this was the case and noted that the precipitation of Schreber's paranoid symptoms, representing a repudiation of homosexual wishes, coincided with the development of a conviction that God was intent on castrating him, and that his movement toward recovery began with a decision to accommodate himself to the female role.

Nunberg (1938) further elucidated the quality of the male paranoiac's attraction to other males. He discerned in his patient a wish to absorb the male object's strength through magical sexual and aggressive means such as sucking and biting the penis. This intense wish to appropriate another male's strength was related to a history of insult to the patient's masculine self-esteem. Consequently, the type of male object most desired was one who was big and strong, but, because of coexistent hostile wishes generated by the insult and jealousy, anticipatory fears of injury from such men necessitated a more gradual approach to them. This was effected through sexual-aggressive contacts with weaker, effeminate-appearing men. After successful contacts with weaker men, the patient felt increasingly powerful and approached stronger-appearing males with increasing confidence and aggressiveness. There was also experienced increased sexual potency with females as a consequence of these contacts with more powerful males. In fantasy this represented the wish and ability to take possession of the mother whom the patient felt could love only a big strong male such as the father had appeared to him as a child.

---

[1] This study is based on a PhD dissertation submitted to the Department of Psychology, University of Michigan and was supported by a predoctoral National Institute of Mental Health research fellowship.

This observation is related to one of the determinants Freud (1950d) found of homosexual object choice in males. Freud stated that the fear of the father was so intense in such cases that, "In his imagination all women belonged to the father, and he sought refuge in men out of submission, so as to 'retire from the conflict' in favor of the father [p. 216]." In line with this, Eidelberg (1945) described a paranoid patient who was able to engage in heterosexual intercourse only when the therapist, playing the role of a strong father figure, encouraged him (i.e., gave him permission) to do so.

Knight (1940) also stressed the paranoid's coexistent hostile and erotic wishes towards the father. However, because of the greater strength of the hostility, approaches to the father figure arouse the anxiety that, "both the object and himself will be destroyed by the hostility in himself and the consequent hostility aroused in the object [p. 156]." This factor is demonstrated in the sequential object choice observed in Nunberg's case.

Ovesey (1955) inferred that intense power struggles between the child and parents and/or older siblings generate extreme hostility which the child could not satisfactorily discharge because of the overwhelming alignment of forces against him. This produced chronic feelings of inadequacy as a male along with the need to be dependent on a perceived omnipotent power figure. These strivings for gratification from a stronger male and the hostile wishes towards him may be expressed through any and all of the sexual modes and zones. Such an individual feels unattractive to women and unable to fulfill other competitive aspects of the male role. This incapacitation, given our cultural value system, contributes to a growing anxiety and conviction that one is homosexual.

Although varying emphasis is placed on the primacy of the libidinal components of the attraction in these accounts, none of these clinicians eschews sexual qualities of the approach motivation altogether. There is a communality of theoretical constructions revolving around the themes of self-perceived power deficits, attraction and aversion to powerful father figures and a secondary avoidance of women.

Psychodynamics and Conflict Theory

Miller's (1944, 1959) approach-avoidance conflict model provides a useful articulated schema for integration with these psychoanalytic conceptualizations of the male paranoid's motivational dynamics in respect to other men. The following is an attempt to portray the postulated dynamics in terms of the Millerian conflict model.

Due to a self-perceived power deficit, the male paranoid is motivated to approach powerful males in order to appropriate their power and its representations (e.g., the penis) by magical sexual and aggressive means. Unlike other types of homosexually motivated males, the paranoid appears extremely unwilling to interact in a consummatory manner with the male(s) to whom he is attracted (i.e., the goal object) to fulfill his wishes. This avoidance tendency is characterized by an anticipated fear of retaliatory punishment, injury and destruction from the male power-figure for his appropriative and aggressive wishes. The intense strength of this avoidance tendency makes the paranoid unwilling to think directly of his wishes for other males and consequently he attempts to alter his own and others' perception of the situation in the various ways described by Freud (1950c).

Fenichel (1954) vividly portrayed the type of motivation, assumed here for the male paranoid, in the following passage.

> I will surrender myself to a great man, so that while doing so I may take away from him his insignium of power, which I want for my own use; I will surrender myself to a man I originally hated, so that I can feel like him and participate in his masculinity; and I will do something to him to compel him to let me participate, if he does not do so of his own accord.
>
> Again, we see, then the same thing: there is an originally aggressive form of seizing from the possessor power his power, in order to possess it oneself. . . .

In sum, it is hypothesized that male paranoids have an intense approach-avoidance conflict, characterized by a need for power and an attendant fear of being destroyed by the powerful male by whom they are aroused.

Before going on to describe an experimental test of these hypotheses, the implications of the model will be further discussed and related to clinical and everyday observations.

Clinical Observations and the Power Hypothesis

The conflict is presumed to result from chronic, severe insult to those aspects of self-esteem that form a central integral portion of the male role as commonly conceived in our culture (e.g., dominance, strength, potency, social adequacy, intellectual competence, competitive adequacy. etc.). These characteristics can be subsumed under the more inclusive term—power. Psychological power can be generally defined as the ability to produce, for oneself or others, nonneutral effects (i.e., effects which represent actual or potential rewards and punishments). Power can be subdivided into the following realms: physical activity (strength), sexual activity (potency), intellectual activity (omniscience), and social activity (influence).

Turning to what is generally regarded as the most nuclear paranoid feature, persecutory delusions, we note that their content portrays the perception of being controlled in these various power realms by a more powerful other (i.e., the persecutor). "Someone is making, or preventing, me from doing things." "Someone is affecting me sexually." "Someone is controlling my thoughts," or "Someone is making others dislike me."

Delusions of grandeur, from this point of view, may be regarded as magical, unrealistic attempts to compensate for the self-perceived power deficit. The grandiose paranoid presumes not only that he is some important figure, as the psychoanalytic theory of narcissistic involvement would have it, but in particular that he is some extraordinarily powerful person in some power realm (e.g., Christ, Hitler, Einstein, etc.).

Power and Sexuality

The relationship of the power deficit to sexual motives poses a difficult problem. Somatically oriented theorists would view passive, receptive sexual wishes as antecedent to self-perceived power deficits (e.g., Freud, Ferenczi and Stärke) while theorists oriented more toward interpersonal relations (e.g., Adler, Sullivan and Ovesey) would see a self-perceived power deficit as antecedent to passive, receptive sexual wishes. Previously cited clinical observations indicated the presence of both sexual and interpersonal components. Instead of prematurely oversimplifying the problem by denying the importance of one or the other component it might be heuristically advantageous to explore how they might be interrelated developmentally.

The concept of sex role provides a conceptual basis for the clarification of this interrelationship. In Western cultures exclusive heterosexual motivation and sexual-object choice are typically prescribed as a nuclear attribute of the sex roles. The successful fulfillment of this prescription for the male role is socially conceived of, and psychologically learned as, particularly demanding of a surfeit of power. Thus, in the realm of sexual behavior, the male is expected to assume a spatially superior position in coitus, often referred to as the dominant position, expressing the power phrasing.

The sexual aspects of the male role are not the only ones which are heavily laden with power meanings. Throughout the life cycle, males are expected to exhibit power in all of the previously mentioned power realms. Furthermore, through the sex-typing of behavior, an expectation of an uneven distribution of power between sex roles, with the male favored, is communicated, encouraged, and reinforced. In contrast the female is seen as more dependent and helpless. Whether this social conceptualization is primarily a reflection of genetic differences in relevant anatomical and physiological drive endowments in respect to power realms is obviously debatable, but relatively unimportant, at this stage of hypothesis construction in psychopathology.

Given this social situation, the male paranoid is theoretically expected to avoid the female as a sexual-object choice because her sexual-social conquest, satisfaction, and maintenance are viewed as sources of demand on a power supply already perceived as limited or depleted. Ordinarily women are attractive to men partially because their sexual-social conquest, maintenance, and satisfaction offer a challenge to the male's power and therefore represent a potential source of enhanced self-esteem. The male paranoid, however, would usually be unwilling to risk the gamble because the anticipated probability of success is so unfavorable.

An exaggeration of an involvement with women based primarily on this power-enhancing factor is not unknown. This kind of involvement has been described as a compensatory attempt to bolster feelings of power depletion. The flight of the aging male, frightened at the prospect of his declining power in many realms (e.g., economic, physical, sexual, etc.) into heterosexual promiscuity is a familiar example. In paranoia this may also be observed in the form of symptom labeled erotomania.

The avoidance of females as sexual objects by male paranoids does not, therefore, imply an absence of heterosexual motivation, but may mean, as in cases described by Nunberg, Eidelberg, and Ovesey, that heterosexual goals must be deferred until the self-perceived power deficit is remedied.

It is of interest to note that Freud (1950b) observed an increase in one of his male patient's paranoid tendencies after the completion of heterosexual intercourse. To encompass this theoretically puzzling phenomenon, Freud invoked the conceptualization of libido in terms of homosexual and heterosexual components which were simultaneously aroused, but which were not simultaneously satisfied. Since intercourse with the female satisfied only the heterosexual component, the homosexual component was left aroused and unsatisfied which accounted for the observed increase in paranoid symptoms.

From the viewpoint of the power-deficit hypothesis, this observation is conceptualized, with greater parsimony, as a manifestation of an increased feeling of power depletion after the completion of a task perceived as power draining.

The cultural view of the sexual act as power-depleting is undoubtedly reinforced by the enormous energy expenditure involved with consequent feelings of fatigue and exhaustion so frequently experienced.

A further testament to the importance of the role of self-perceived power depletion in the dynamics of male paranoid symptomatology is well illustrated by the reported precipitating circumstances of paranoid episodes in the published case histories. The individual has usually suffered or anticipated an insult to some aspect of his male role heavily laden with power meaning. The resulting humiliation primarily reflects on the adequacy of his power (e.g., infidelity of his wife, business failure, inability to manage subordinates, failure in athletic competition, etc.). Since these impressions were recorded by various clinicians independently of the formulation or entertainment of a power hypothesis, they constitute impressive confirmatory clinical evidence.

It has been posited that the power deficit resulted from the effects of early prolonged conflict with more powerful males. If, however, a female played a major role in the power conflict, it might be expected that a female would substitute in the role depicted for a male in symptom dynamics. If a female played an adjunctive role in the genesis of the power conflict, both sexes will be represented in symptomatology.

Assuming that a male played a major part in the conflict genesis, it might still be expected that the female, through stimulus generalization, would cue off the power conflict. That is, paranoid males would be as attracted and repelled by powerful appearing women as they would by powerful appearing males. However, this does not appear to be the case! Clinical observation of male paranoid patients has long resulted in the impression that the persecutor is almost exclusively male. This impression was recently confirmed in a survey of male paranoids' persecutory delusions conducted by Klaf and Davis (1960) who found the persecutor to be exclusively male in about 85% of the cases as opposed to female in about 5% of the cases. In the remaining 10% of the cases persecutors were male and female.

The reason why generalization fails to occur may be understood in terms of sex-typing experiences which act much like discrimination training in that only like-sexed persons tend to be seen as models, and hence as suitable objects from whom power may be obtained. Thus even when the paranoid male occasionally encounters powerful females they do not have the same effect as encounters with powerful males.

Because of the effects of sex-typing in our culture, power and male sexuality are closely interwoven. Consequently, behavior which appears to be sexually motivated may also have power components. This theory has suggested that the wish for power is an important determinant of the male paranoid's attraction towards, and fear of, other males. The relationship of this intense approach-avoidance conflict to a variety of paranoid symptoms has been outlined.

METHODOLOGICAL ASSUMPTIONS AND RATIONALE

S-R conflict theory assumes that the strength of the approach or avoidance tendency changes as a function of the distance between person and goal object. Such assumptions are not foreign to psychoanalytic thinking. Fenichel (1945) has written, "The first positive instinctual behavior toward a desired object consists in diminishing the distance between oneself and the object. . . . The first negative instinctual behavior toward a repulsive object consists in increasing the distance [p. 83]."

One further relevant assumption of conflict theory is that approach and avoidant tendencies interact in such a way that the actual distance between a person and an object, given determination by the person, represents an algebraic resultant of the two forces.

Experimental support for the relationship between distance and attraction is reported in studies by Smith (1952) and Shellow (1956).

A more concise statement of the theoretical hypothesis and its implications is necessary. As the perceived power of other males is increased, male paranoids experience an increase in the strength of both their approach and avoidance tendencies. These changes, based on the assumption that the increment is equal in both types of tendencies, are depicted in Figure 1.

Increments in the strength of the approach tendency are represented by the changes from 1 to *2* to 3, while increments in the strength of the avoidance gradient are represented by the changes from I to II to III. The effects of these simultaneous increments in the strength of both tendencies on the equilibrium point can be followed in the successive changes of the ordinate and abscissa coordinates from A to B to C.

The strength of the conflict increases, as indicated by the successive increments in the ordinate values of the equilibrium point, and the *resultant* attraction is diminished as indicated by the successive increments in the absicissa values. This latter effect is the major deduction from the theory which will be subject to experimental test. A restatement of the hypothesis follows: Male paranoids react to increments in the perceived power of other males with decreased resultant attraction to them (as reflected in increased distance between the objects and themselves), but are not similarly affected by increments in the perceived power of females.

METHOD

To measure resultant attraction, the dependent variable, an apparatus was designed which allowed the subjects to manipulate the distance between themselves and photographs of people and inanimate objects. The apparatus consisted of an 8-foot enclosed tunnel in which a moveable platform, illum-

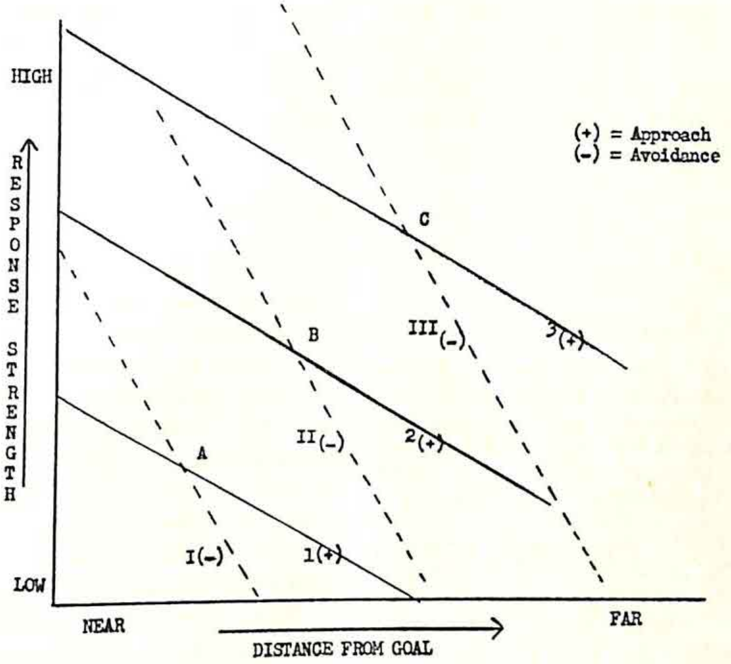


FIG. 1. Approach-avoidance conflict model depicting the effects of equal increments in the strength of approach and avoidance gradients on distance.

inated by panel lights only, was controlled by cranks located at either end of the tunnel. This enabled the experimenter at one end, or the subject at the other end, to move the photograph. The subject was instructed to look through a small opening at his end, into the tunnel, and to move the picture, placed there by the experimenter, towards him until he found the place where "it looked best." The subject to photograph distance was indicated to the experimenter on a tape measure by a marker on the outside of the tunnel coordinated with the moveable platform.

The instructions were designed to allow the subject to be influenced by his approach and avoidance tendencies without making him conscious of their role in the determination of photograph placement.

Both to establish a neutral set and to control for visual acuity, the subject first manipulated a standard reading card to various positions in the tunnel in order to read the varying sized print on the card. After this familiarization with the apparatus and measure of visual acuity the experimenter presented the photographs in randomized serial order.

Photographs of people's faces were used as the source of power cues because the power connotations involved in facial features and expression depend on relatively subtle visual cues that can be easily obliterated by small increments in the optimal viewing distance between the subject and object. Furthermore, such stimuli are not likely to arouse overly disturbing and interfering feelings in the subject.

To avoid contaminating the subjects to be used to test the hypothesis by arousing their defenses, the ratings of perceived power were obtained from 64 volunteer college students of both sexes. One hundred and fifty-three photographs were selected to be rated from the hundreds examined. All were achromatic, nonglossy, above the shoulder detail, contour cut, and presented on black backgrounds. Each student rated about 50 photographs on Osgood's (Osgood, Suci, & Tannenbaum, 1957) Semantic Differential since one of the three major factors of meaning measured, potency, approximates what we have referred to as power. Thus the potency factor loads highly on such bipolar adjective scales as strong-weak, hard-soft, large-small, and was appropriately nicknamed the "football player" factor. Osgood reports that the potency factor when combined with the other two major factors, evaluation and activity, accounted for much of the variance in the discriminable meaning of objects and concepts for subjects of all types. Thus the value of a photograph on these three factors could be expected to account for much of the psychological impact which it had on subjects. Experimental or statistical control of the photographs on the evaluation and activity factors, which are extraneous to the present study, would allow us to conclude that differential reactions to the photographs were due to the remaining potency factor.

The mean value of each photograph on each of the three Semantic Differential factors was computed. An unsuccessful attempt was made to obtain photographs which would differ only on potency values. Since experimental control of the values of these extraneous variables was not possible, the control of their effect and that of photograph size was deferred to statistical techniques for partialing them out.

Twenty-four male and 19 female photographs were selected to which were added 5-unrated photographs of inanimate neutral objects including: a chair, building, table, car, and ten-of-spades playing card. This brought the total number of photographs to 48.

Operational Hypotheses

We are now enabled to summarize our previous theoretical discussion in operational terms. The initial psychoanalytic conceptualization of the relationship between the sex of the goal object and the male paranoid's attraction-repulsion was supplemented by a more differentiated conceptualization emphasizing the perceived power of a like-sexed person as the basis for the responses. Henceforth, this will be referred to as the power-sex hypothesis and can be stated as follows: As rated potency values of male photographs increase, male paranoids place them further away from themselves in contrast to their distance placements of female photographs which are not affected by increases in their rated potency values.

If, however, the older simple homosexual hypothesis is entertained, male paranoids would place photographs of males closer than either photographs of females or of neutral objects, and would be likely to place the neutral photographs closer than the female photographs.

Subjects

All subjects in the study were hospitalized male schizophrenics who had previously served in a comprehensive biochemical study of schizophrenia independent of the present study. Of the 67 subjects chosen for inclusion in the experiment, 5 refused to cooperate, 2 were discarded because they correctly inferred the purpose of the task, and 1 was unable to follow the instructions, thus reducing the number of participating subjects to 59. Five of the 8 subjects discarded were paranoid and 3 were nonparanoid.

The experimenter had access to five separate psychiatric diagnoses for each subject. Notes on the diagnostic interviews by the psychiatrists indicated they were more likely to erroneously diagnose a patient as nonparanoid than as paranoid. Therefore a liberal classification rule was employed such that if a subject was diagnosed by any one of the five psychiatrists as paranoid schizophrenic he was included in the paranoid group. This procedure yielded 35 paranoid schizophrenics and 24 nonparanoid schizophrenic control subjects.

In 72% of these cases at least three of the five psychiatrists agreed on the paranoid subclassification. The frequency distributions of the two groups were compared for age, education, religion, skin color, veterans status, intelligence, length of hospitalization, visual acuity and ataraxic drug medication. No significant differences between the two groups were found on these variables although the paranoid groups tended to have somewhat higher intelligence.

RESULTS AND DISCUSSION

Power-Sex Hypothesis

Two subjects may place a given photo at equal distances from themselves, but for one subject this may be his closest placement, whereas for the other subject it may be his farthest placement. To circumvent this ambiguity involved in using absolute distance values as an index of the attractiveness of the photo for the subject, each of the 48 photos was ranked for each subject on the basis of its relative distance from the subject in comparison with his other photo placements.

The 35 rank scores obtained for each photo from the paranoid group and the 24 rank scores obtained from the nonparanoid group for each photo were averaged yielding a group mean for each photo.

Since the photos differed in terms of their values on the extraneous variables of evaluation, activity and size, which threatened to confound the effects of variations in their potency values on their placement, the effects were removed statistically through the method of partial correlation. These results are presented in Table 1.

The third order partial correlation ($r = 0.408$) indicates that the greater the potency value of male photos the farther away the paranoids place them whereas the nonparanoid schizophrenics' placements are relatively unaffected by the potency values of male photos ($r = 0.107$).

In addition to the difference between para-

TABLE 1

THIRD ORDER PARTIAL CORRELATIONS BETWEEN DISTANCE SETTINGS AND POTENCY VALUES OF PHOTOGRAPHS

	Sex of photograph	
	Male ($N = 24$)	Female ($N = 19$)
Paranoid schizophrenics	0.408	0.019
Nonparanoid schizophrenics	0.107	0.579

noid and nonparanoid schizophrenics in their reactions to potency values of male photos, the hypothesis also predicted that paranoids' placements of the female photos would be unaffected by their potency values. The correlation ($r = 0.019$) indicates this to be the case.

Although the hypothesis did not specifically deal with the nonparanoid schizophrenics' reaction to potency in females, it is of interest to note their considerable tendency ($r = 0.579$) to place photos of potent looking women farther away in contrast to their relative lack of reactivity to potent looking males.

Since the hypothesis described a difference within the paranoid group in response to the two types of photos as well as a difference between the paranoids and nonparanoids, a single test of significance does not suffice. However, the probability values derived from each of the two relevant significance tests for differences between third order partial correlations (Walker & Lev, 1953) may be combined (Fisher, 1946) to yield a probability estimate of the joint occurrence of both differences relevant to the hypothesis. This estimate of the probability of the joint occurrence of the between-subject group differences on the male photos and the within-paranoid group difference on the male versus female photos is $.10 > p > .05$, 1-tailed.

Undifferentiated Homosexual Hypothesis

Since the traditional psychoanalytic viewpoint attributes paranoid conflict concerning homosexual object choice solely to the sexual identity of the object without further specification of the attributes of attraction and repulsion within like-sex type, it was referred to as an undifferentiated homosexual hypothesis. Results reflecting on this hypothesis were obtained through the following procedure: Each subject's mean distance rank for each of the three photo types (i.e., male, female, neutral object) was calculated. Then, for each subject, the three differences between these means were computed, (i.e., the mean for the male photos minus the mean for the female photos, etc.). The mean of these three distributions of differences was found for the paranoid and nonparanoid groups separately and are presented in Table 2. In Table 2 a positive value for a mean difference indicates that the former of the two photo types being compared was placed farther from the subjects than the latter type. Thus both the paranoids and nonparanoids placed male photos farther from themselves than photos of females, etc. The critical question though is whether paranoids tend to do so significantly more than nonparanoids.

The results in Table 2 indicate that of the three differences relevant to the undifferentiated homosexual hypothesis only one is consistent with it. The paranoids place photos of females significantly farther away from themselves than neutral photos as compared to the nonparanoids ($p = .05$, 2-tailed).

The other results in Table 2 indicate that: (*a*) Paranoids placed male photos farther away from themselves than female photos as compared to nonparanoids. However this greater tendency on the part of the paranoids to place male photos farther from themselves than female photos was not significant ($.20 > p > .10$, 2-tailed) as were the other differences. (*b*) Paranoids placed male photos significantly farther from themselves than photos of neutral objects as compared to nonparanoids ($p < .01$, 2-tailed) who placed both

TABLE 2

DIFFERENCES BETWEEN MEAN RANK DISTANCES OF MALE, FEMALE, AND NEUTRAL PHOTOGRAPHS

	Mean difference		
	Male–female photographs	Male–neutral photographs	Female–neutral photographs
Paranoid schizophrenics ($N = 35$)	3.03	6.09	2.99
Nonparanoid schizophrenics ($N = 24$)	1.45	0.08	−1.36
"t" (df-57)	1.31	2.86	1.99
p	$.20 > p > .10$, 2-tailed	$p < .01$, 2-tailed	$p = .05$, 2-tailed

types of photos at approximately equal distance from themselves.

It was unlikely that the subjects' responses were guided by cues unwittingly emitted by the experimenter because: (*a*) discussions were limited to the "cover" task set—visual acuity, (*b*) subjects did not look up from the tunnel until they signaled they had finished the trial, (*c*) the experimenter was ignorant of the photo content on any particular trial because they were randomized and face-down, (*d*) subjects were questioned after the experiment to determine possible detection of its real purpose.

These results are more consistent with the pattern of specific differences described in the power-sex hypothesis which was derived from a theoretical interpretation of paranoid symptomatology based on an attraction and repulsion in respect to power in like-sexed persons (other males in this case). This notion is a supplement to the traditional psychoanalytic conceptualization of the approach-avoidance conflict experienced by male paranoids towards other males. Expectations generated on the basis of the traditional hypothesis were not as consistent with the data of this study.

In contrast to the paranoids, nonparanoid schizophrenics' intense reaction to power in women was not particularly anticipated since their relationship to women was not the focus of this study. It suggests, however, that their experience with significant females has been different. Since nonparanoid schizophrenics show decreased attraction to increased power in females, it specifically suggests that the mother was perceived as a traumatizing figure parallel to the perception of the father for paranoids. Because of the reciprocal nature of male-female roles, severe traumatization experienced by the child from a parent of either sex would be expected to generate the severe kind of sexual identity conflicts observed in all types of schizophrenia. However if, as suggested by the data of this study, there is a difference between paranoid schizophrenics and nonparanoid schizophrenics in the sexual identity of that parent, or significant other, perceived as the major traumatic figure certain qualitative differences in symptomatology ought to ensue.

The intense pressures involved in the sex typing of children's behavior generates strong needs towards assuming those particular characteristics the cultural tradition prescribes for a particular anatomical sexual identity. Resulting conformity motives are so intense that even most effeminate males usually conform to most of the attributes associated with maleness in our culture (e.g., hair style, dress, absence of facial make-up, etc.). Power was previously described as such a "sex-linked" attribute in Western European cultures. Men are supposed to have a lot of it and to cultivate it, while women are not. This socialization pressure towards identification with the attributes prescribed for one's anatomical sex type is undoubtedly one of the factors responsible for the strength of the approach gradient in male paranoids towards the perceived powerful traumatic father. Such an identificatory motive would be relatively lacking in the male nonparanoid schizophrenic's reaction to the inferred traumatic powerful mother. Consequently, for nonparanoid schizophrenic males, while the avoidance tendency regarding powerful females might be strong, the approach tendency is likely to be weak compared to the strength of the approach tendency of paranoid males toward powerful males. Therefore the nonparanoids would not be involved in the intense approach-avoidance conflict that the paranoids are beset by. Nonparanoid schizophrenic males would be expected to avoid powerful women without the telltale expression of overconcerned interest in males that characterizes the male paranoid's symptom picture. Thus, we find an absence of persecutory delusions concerning women in male nonparanoid schizophrenics.

It might appear that this is tautological since the appearance of any strong persecutory theme in a schizophrenic would likely result in his being classified as paranoid. However, the previously cited survey of persecutory symptoms among male paranoids (Klaf & Davis, 1960) showed that only 5% of the male paranoid schizophrenics imagined a female solely as persecutor in contrast to 85% who imagined a male solely as persecutor.

Research on the schizophrenogenic mother has generally made little attempt to relate results to subtype of schizophrenic diagnosis.

TABLE 3

PARENTAL DOMINANCE BY SCHIZOPHRENIC SUBCLASSIFICATION OF SON

	Indices showing mother dominant or equal	Indices showing father dominant	Total
Paranoid schizophrenics ($N = 13$)	31	47	78
Nonparanoid schizophrenics ($N = 9$)	38	16	54
Total	69	63	132
	$\chi^2 = 13.0$	$p < .001$	

Note.—$\chi^2 = 13.0$, $p < .001$.

If, as our results suggest, paranoid and nonparanoid schizophrenics' experiences and perceptions of their parents differ, inconsistent results of investigators may be due to unreported differences in schizophrenic subtype sample composition. Absence of the relevant information generally precluded possible verification of this hypothesis.

In the one attempt made, however, the results were quite encouraging. Farina[2] generously supplied his subclass data on the sample of schizophrenics he employed in a study of parental dominance through the method of revealed differences in decision making. Farina's (1960) focus was on the quality of premorbid adjustment as related to parental dominance. On the basis of the findings of the present study, it was expected that there would be a significantly greater number of indices showing father dominance in the male paranoid schizophrenic group ($N = 13$) than in the male nonparanoid schizophrenic group ($N = 9$). Six of the interaction measures available for the parents of each subject provided evidence of parental dominance. The distributions of these indices are reported in Table 3.

Again the data from this independent sample support the hypothesis that the fathers and mothers of male paranoid schizophrenics significantly differ from those of male nonparanoid schizophrenics in terms of power and dominance—the fathers being significantly more dominant than the mothers in the case of paranoids.

The applicability of the notions forwarded in this study to female paranoids is uncertain and must be carefully studied.

---

[2] A. Farina is not responsible for the particular conclusion drawn here from his data.

REFERENCES

ARONSON, M. L. *A study of the freudian theory of paranoia by means of a group of psychological tests.* (Doctoral Dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1950, No. 2374.

BRILL, A. A. Homoeroticism and paranoia. *American Journal of Psychiatry,* 1934, **90**, 957–996.

DASTON, P. G. *Perception of homosexual words in paranoid schizophrenia.* (Doctoral Dissertation, Michigan State College) Ann Arbor, Mich.: University Microfilms, 1952. No. 4691.

DEVINE, H. The biological significance of delusions. *Journal of Mental Science,* 1916, **62**, 135–150.

DEVINE, H. A study of hallucinations in a case of schizophrenia. *Journal of Mental Science,* 1921, **67**, 172–186.

EIDELBERG, L. Psychoanalysis of a case of paranoia. *Psychoanalytic Review,* 1945, **32**, 373–402.

FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1960, **61**, 31–38.

FENICHEL, O. *Psychoanalytic theory of neurosis.* New York: Norton, 1945.

FENICHEL, O. Trophy and triumph (1939). *Collected Papers of Otto Fenichel.* Series 2. New York: Norton, 1954.

FERENCZI, S. The nosology of male homosexuality (1914). In E. Jones (Ed.), *Sex in psychoanalysis.* New York: Brunner, 1950. (a)

FERENCZI, S. On the part played by homosexuality in the pathogenesis of paranoia (1912). In E. Jones (Ed.), *Sex in psychoanalysis.* New York: Brunner, 1950. (b)

FERENCZI, S. Some clinical observations on paranoia and paraphrenia (1914). In E. Jones (Ed.), *Sex in psychoanalysis.* New York: Brunner, 1950. (c)

FERENCZI, S. Stimulation of the anal erotogenic zone as a precipitating factor in paranoia (1911). *Selected Papers of Sandor Ferenczi.* New York: Basic Books, 1955.

FISHER, R. A. *Statistical methods for research workers.* (10th ed.) London: Oliver and Boyd, 1946.

FREUD, S. A case of paranoia running counter to the psychoanalytic theory of the disease. (Orig. publ. 1915). In, *Collected papers.* Vol. 2. London: Hogarth, 1950. (a)

FREUD, S. A neurosis of demoniacal possession in the seventeenth century. (Orig. publ. 1923). In, *Collected papers.* Vol. 4. London: Hogarth, 1950. (b)

FREUD, S. Psychoanalytic notes upon an autobiographical account of a case of paranoia. (Orig. publ. 1911). In, *Collected papers*. Vol. 3. London: Hogarth, 1950. (c)

FREUD, S. The psychogenesis of a case of homosexuality in a woman. (Orig. publ. 1920). In, *Collected papers*. Vol. 2. London: Hogarth, 1950. (d)

FRINK, H. W. The significance of a delusion. *Psychoanalytic Review,* 1926, 13, 16–31.

GARDNER, G. E. Evidence of homosexuality in 120 unanalyzed cases with paranoid content. *Psychoanalytic Review,* 1931, **18**, 57–62.

KLAF, F. S., & DAVIS, C. A. Homosexuality and paranoid schizophrenia: A survey of 150 cases and controls. *American Journal of Psychiatry,* 1960, **116**, 1070–1075.

KNIGHT, R. P. The relationship of latent homosexuality to the mechanism of paranoid delusions. *Bulletin of the Menninger Clinic,* 1940, **4**, 149–159.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

MOORE, R. A., & SELZER, M. L. Male homosexuality, paranoia and the schizophrenias. *American Journal of Psychiatry,* 1963, **119**, 743–747.

MUSIKER, H. R. Sex identification and other aspects of the personality of the male paranoid schizophrenic. Unpublished doctoral dissertation, Boston University, 1952.

NORMAN, J. P. Evidence and clinical significance of homosexuality in 100 unanalyzed cases of dementia praecox. *Journal of Nervous and Mental Diseases,* 1948, **107**, 484–489.

NUNBERG, H. Homosexuality, magic, and aggression. *International Journal of Psychoanalysis,* 1938, **19**, 1–16.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

OVESEY, L. Pseudohomosexuality, the paranoid mechanism and paranoia. *Psychiatry,* 1955, **18**, 163–173.

PAGE, J., & WARKENTIN, J. Masculinity and paranoia. *Journal of Abnormal and Social Psychology,* 1938, **33**, 527–531.

SHELLOW, R. S. *Perceptual distortion in the spatial localization of emotionally meaningful material.* (Doctoral Dissertation, University of Michigan). Ann Arbor, Mich.: University Microfilms, 1956, No. 18649.

SHOCKLEY, F. M. The role of homosexuality in the genesis of paranoid conditions. *Psychoanalytic Review,* 1914, **1**, 431. (Abstract)

SMITH, G. H. Size-distance judgments of human faces (projected images). In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N.H.: Institute of Advanced Research, 1952.

STÄRCKE, A. The reversal of the libido-sign in delusions of persecution. *International Journal of Psychoanalysis,* 1920, **1**, 231–234.

VON OPHUIJSEN, J. H. On the origin of the feeling of persecution. *International Journal of Psychoanalysis,* 1920, **1**, 235–239.

WALKER, HELEN, & J. LEV. *Statistical inference.* New York: Holt, 1953.

ZAMANSKY, H. S. An investigation of the psychoanalytic theory of paranoid delusions. Unpublished doctoral dissertation, Harvard University, 1953.

(Received October 9, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 371–377

NEED FOR STIMULATION AS A SOURCE OF STRESS RESPONSE TO PERCEPTUAL ISOLATION[1]

MARVIN ZUCKERMAN AND MERRY MARQUIT HABER[2]

Research Laboratories, Division Endocrinology & Reproduction, Albert Einstein Medical Center, Philadelphia

Adelphi College

The study was undertaken to see if Ss who had shown a greater stress reaction (GSR reactivity) to perceptual isolation could be shown to have a greater "need" for stimulation than those who were not so stressed by isolation. Ss selected on the basis of their high or low GSR reactions to the prior isolation experiment were tested in a second 3-hr. perceptual isolation situation, only this time they were given the opportunity to make an operant response which would produce random visual or auditory stimulation depending on their choice. Those previously stressed by isolation made significantly more responses for visual and auditory reinforcement than the low-stress group. All Ss responded more for visual than for auditory reinforcement.

An examination of the natural behavior of animals or humans shows a remarkable quantity of behavior aimed at no obvious avoidant or appetitive need reduction. In fact, much of this behavior seems to have as its end the increase or variation of stimulation rather than its reduction. In the last 10–15 years psychologists have attempted to study these phenomena in the laboratory.

Baron and Kish (1962) found that low-intensity auditory and visual stimuli can reinforce operant response in the mouse. Moon and Lodahl (1956) have demonstrated that changes in intensity of light stimulation in either direction will reinforce lever pressing in the rhesus monkey. A series of experiments by Butler (1953, 1954, 1957 and 1958) and others (Butler & Harlow, 1954; Fox, 1962; Harlow & Zimmerman, 1959; Kish, 1955; Levine & Forgays, 1959; Myers & Miller, 1954; Premack, Collier & Roberts, 1957; Robinson, 1957) have demonstrated that visual, auditory, or tactual stimulation may serve as reinforcement without any obvious association with primary reinforcement. There is much in the data which suggests that stimulus deprivation arouses a primary "drive." The behavior is remarkably resistent to extinction or satiation (Butler & Harlow, 1954). Butler and Alexander (1955) found no consistent trends in a window opening response over daily trials. Monkeys tested for 10 hours on 6 consecutive days spent about 40% of their time looking out into the laboratory suggesting that animals worked to maintain some "relatively fixed amount of daily visual experience." Fox (1962) found increased operant response with increasing deprivation time up to 4 hours. The nature of the stimulus may be crucial. Butler (1954) compared the reinforcing power of the sights of an empty room, food, a toy-train in motion, and another monkey, and found them from lowest to highest in the order given above. The low rank of food in the hierarchy is interesting in its contradiction of the secondary reinforcement explanation of these phenomena.

The new research area of sensory deprivation gave some impetus to the study of stimulation need in man. Jones (1961, 1964) and Jones, Wilkinson, and Braden (1961) found that visually deprived subjects responded for light flashes as reinforcement. A drive interpretation for information deprivation was supported by increasing initial response rates as a direct linear function of length of prior deprivation up to 48 hours, the limits of these experiments. The most important factor in the stimulus was its information value, that

---

[1] This research was supported in whole by Public Health Research Grant MH-07926-01. The authors wish to acknowledge the generous assistance of D. Vincent Biase and Sheldon Levine.

This paper was based on a thesis submitted by the junior author, in partial fulfillment of the requirements for the PhD degree at Adelphi College, under the direction of the senior author.

[2] Now at Albany Study Center for Learning Disabilities, Albany, New York.

is, greater randomness or unpredictability in the series was associated with greater operant responding.

A review of "perceptual isolation as a stress situation" by Zuckerman (1964) indicates that sensory deprivation is stressful for many subjects as indicated by much of the verbal, behavioral, and physiological data. Furthermore, the reported visual and auditory sensations (Zuckerman & Cohen, 1964b) and the attempts at self-stimulation suggest that the subjects are attempting to cope with a deficit in exteroceptive stimulation.

The hypothesis which suggests itself is that the individual differences in stress reactions may be a function of differences in stimulation need. Supporting this hypothesis is Vernon and McGills' (1960) experiment. Subjects in this study were given the opportunity to push a button illuminating a viewing box containing a drawing of a line and two circles. Nine subjects who completed 72 hours of deprivation were compared with 6 quitters on their viewing time during the first 24 hours of sensory deprivation. The quitters spent significantly more time looking into the box than the stayers. Another relevant finding comes from Hull and Zubek's (1962) comparisons of stayers and quitters in sensory deprivation on a biographical inventory: quitters watched television more than stayers. These findings would suggest that persons with a greater need for extroceptive stimulation will be more stressed by the deprivation of such stimulation than others.

Zuckerman, Levine, and Biase (1964) tested 36 female undergraduates in three groups: (*a*) total isolation, no light or sound, (*b*) auditory deprivation, light but no sound, and (*c*) visual deprivation, sound (music) but no light. All subjects were given 3 hours of total or partial perceptual isolation and continuous GSR recordings were made during that time. There was little difference between groups for the first hour and a half, but during the second hour and a half basal skin conductance tended to rise more rapidly in the total isolation group and the number of spontaneous GSR fluctuations in this group was significantly greater than in the partial isolation groups. Subjects were rated on their responses to a postisolation interview. High GSR responders in the total isolation group complained significantly more about the inactivity and sensory isolation than the low GSR responders.

The hypothesis of the current experiment is that persons who have responded with greater stress to a previous isolation situation will demonstrate higher response rates for visual and auditory reinforcement in a subsequent experiment than will nonstressed subjects from the previous experiment.

METHOD AND PROCEDURE

Subjects

Six weeks after the experiment described above (Zuckerman, Levine, & Biase, 1964) the subjects in each of the three groups were divided into the highest and lowest GSR responders on the basis of the magnitude of their conductance increase during the last hour and a half of the experiment. Conductance change, rather than number of nonspecific GSRs (NS-GSR), was used because there was no significant group difference on this measure. Use of the NS-GSR measure would have precluded obtaining high and low GSR responders from the prior groups who were in comparable ranges. All high GSR responders had increases in conductance of 20 microhmos or more; all low GSR responders had increases of less than 3, or decreases in conductance. Twenty-four subjects were obtained who consented to participate in a second perceptual isolation experience. Table 1 shows the design of the study.

In the previous experiment, three groups were each given a different type of perceptual restriction. The four highest GSR reactors and the four lowest reactors within each group were selected for the current experiment. This resulted in a total of 12 high and 12 low GSR reactors to the previous experiment. Eight subjects (4 high and 4 low reactors) were drawn from each of the prior condition groups. Although our main interest was in the difference between high and low GSR reactors on their performance in the current experiment, the influence of the prior conditions was also analyzed. To these two primary sources of variance was added a third

TABLE 1

DESIGN FOR THE EXPERIMENT

Type of prior isolation restriction	GSR reaction to prior isolation		
	High conductance increase	Low conductance increase or a decrease	Total subjects
Visual and auditory	4	4	8
Visual	4	4	8
Auditory	4	4	8
Total subjects	12	12	24

source of variance: the type of reinforcement (visual versus auditory) reacted to in the current experiment.

Reinforcement Stimuli

The experimenters attempted to create visual and auditory reinforcers with little intrinsic meaningfulness but with high information, or randomness. Meaningful stimuli such as art slides or music would introduce the factor of individual tastes, but Jones (1964) has demonstrated that meaningless stimuli with high information value will sustain responding over long periods of isolation. The visual stimuli consisted of strips of colors painted across a blank film strip with the order and choice of colors selected from a table of random numbers. The auditory stimuli consisted of tape-recorded tones played on a vibraphone, the choice and order of notes being dictated by the table of random numbers. The notes were played and recorded on the tape at $\frac{3}{4}$-second intervals.

In order to test the psychological equivalence of the two kinds of stimuli a set of 14 rating scales, from Osgood's Semantic Differential (1957) research were used. Six scales measured the evaluation factor, 4 scales the potency factor, and 4 scales the activity factor. Such scales have been used in prior studies to measure reactions to sensory stimuli and esthetic preferences (Osgood et al., 1957). A pilot group of 10 subjects viewed the color and heard the tone sequences and rated each set on a separate set of scales. The ratings indicated that the colors were rated more potent than the tones. To compensate this inequivalence the amplitude of the tones was increased for the experimental subjects. As will be seen, the experimental subjects rated the colors and tones as equivalent on the three factors at the outset of the experiment.

Apparatus

The subject was permitted to operate a button which was connected to Foringer relays, a count programmer and counters located outside of the soundproof room. The subject could operate another switch which controlled the choice of reinforcement. In the right position a filmstrip projector was activated, in the left position a speaker in the soundproof room was activated, and in the middle position neither was activated, although the subject might still respond and the responses would be counted. The count programmer was set so that 10 responses by the subject would either advance the filmstrip and open the shutter on the projector for 15 seconds, or would activate the speaker allowing the subject to hear 15 seconds of the random tones on the continuously moving tape. The subject also had the option of responding without visual or auditory reinforcement by leaving the selection switch in the middle position. After the 15-second period the shutter on the slide projector would close, or the speaker would be cut off, and another 10 responses would be necessary to activate them. Timers controlling the counters switched the count to a new counter each hour of the experiment.

The subject was isolated in a 7 × 7 feet double-wall soundproof room (interior dimensions) manufactured by Industrial Acoustics Corporation. When the shutter on the projector was closed the room was totally dark.

Procedure

The subject was given the Affect Adjective Check List (AACL) measure of anxiety state (Zuckerman, 1960) and the Somatic Symptom Check List (SSCL), both of which have been used in previous isolation experiments (Zuckerman et al., 1962, Zuckerman & Cohen, 1964a; Zuckerman, Levine, & Biase, 1964). These are given before the experiment to establish the baseline for change in the same measures given after isolation. In the measures given after isolation the subjects were asked to describe their experiences in the isolation room.

The subject was then conducted into the soundproof room and given the usual instructions about keeping movements to a minimum and not talking. After this, the use of the button and switch were explained to the subject. The subject was allowed to sample the visual and auditory stimuli. Following each demonstration the subject rated each type of stimulus on a separate Semantic Differential scale to obtain baselines for the evaluation, potency and activity "meanings" of the stimuli. Following this, the door of the room was closed and the subject remained in the dark soundproof environment for 3 hours during which time she was free to use the operant button to bring stimulation into the environment. Instructions were given which stated that the use of the button was entirely up to the subject, that this was not an endurance test or a test of "will-power." Prior pilot studies have indicated that some subjects feel that they should restrain themselves, while others feel we are testing their response rates. The experimenter must be careful to attempt to neutralize such cognitive sets in a human operant situation of this type.

After the 3 hours of isolation were over the subject was brought out of the room, given the Semantic Differential to rate each of the types of stimuli and given the AACL and SSCL to describe the period when she was isolated.

RESULTS

Table 2 shows the mean number of responses for visual and auditory reinforcement for the high and low GSR reactors. Table 3 presents the analysis of variance based on the data in Table 2, further subdivided by prior condition of isolation. The difference between the high and low GSR reactors was highly significant. The high GSR reactors in the prior experiment made almost four times as many responses as the low GSR reactors. The dif-

TABLE 2

RESPONSE MEANS OF HIGH AND LOW GSR REACTORS FOR VISUAL AND AUDITORY REINFORCEMENT

Type of reinforcement	Reaction to prior isolation		
	High GSR	Low GSR	All subjects
Auditory	938	239	588
Visual	2,114	594	1,354
Both	3,052	833	1,942

TABLE 4

RESPONSE MEANS OF PRIOR CONDITION GROUPS FOR VISUAL AND AUDITORY REINFORCEMENT

Type of reinforcement	Type of prior isolation restriction		
	Auditory	Visual	Visual and auditory
Auditory	578	663	525
Visual	978	1,281	1,944

ference was particularly striking in responses for the visual reinforcement, where only 1 of 12 High GSR responders made less than 1,000 responses and only 1 of 12 low GSR responders made more than 1,000 responses.

A second highly significant source of variance was the type of reinforcement. All groups made significantly more responses for the visual than for the auditory reinforcer. None of the subjects responded with the switch in the middle position which would have produced neither of the reinforcements.

A significant interaction between the GSR reactor groups and the type of reinforcement was obtained. The difference between the high and low GSR reactors was greater for the visual reinforcement than for the auditory reinforcement.

There was a tendency toward a significant interaction between type of prior isolation and type of reinforcement ($p < .10$). This is shown in Table 4 containing the means for the prior isolation type groups for the light and sound reinforcers. The prior condition of isolation does not seem to have affected the responses for auditory reinforcement, but some effect is discernible in the response for the visual reinforcement. The group which was visually and auditorily deprived made the greatest number of responses, the group which was visually deprived made the next greatest number, and the group which was auditorily deprived made the least number of responses for visual reinforcement.

Figure 1 shows the mean rates of response for the two types of stimuli during the three 1-hour periods of the experiment. Analyses of variance indicated that the time variable was significant for the visually reinforced responses and not for the auditorily reinforced responses. As can be seen in Figure 1, there was no change in responses for the visual reinforcement during the first 2 hours of the experiment, but there was a sharp rise in responding during the third hour of the experiment.

TABLE 3

ANALYSIS OF VARIANCE OF RESPONSES FOR AUDITORY AND VISUAL REINFORCEMENT

Source	*MS*	*df*	*F*
Between subjects		23	
GSR Reactivity (A)	14,763,008	1	27.5***
Type Prior Isolation (B)	598,694	2	1.1
A × B	547,977	2	1.1
Error between groups	532,204	18	
Within subjects		24	
Reinforcement (C) Visual versus Auditory	7,038,008	1	25.6***
A × C	2,025,408	1	7.4**
B × C	837,102	2	3.0*
A × B × C	103,177	2	.4
Error within subjects	274,518	18	

* $p < .10$.
** $p < .05$.
*** $p < .001$.

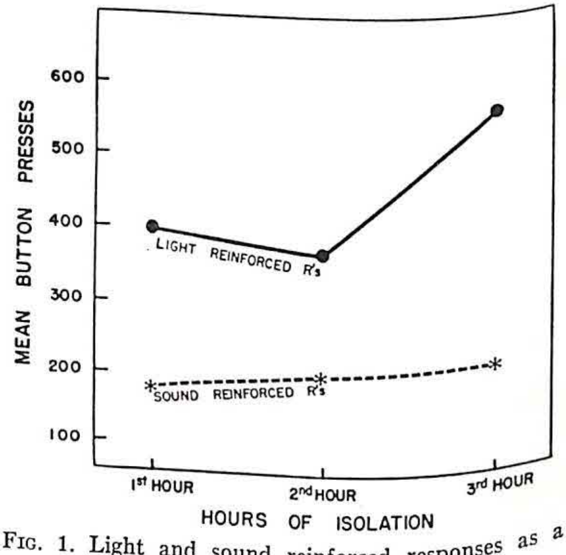


FIG. 1. Light and sound reinforced responses as a function of hours of isolation.

TABLE 5

MEAN POTENCY RATINGS OF VISUAL AND AUDITORY STIMULI

Type of reinforcement stimuli	Before isolation	After isolation
Auditory	3.9	3.5
Visual	3.9	4.2

Semantic Differential Ratings

Analyses of variance indicated no influence of GSR reactivity differences between stimuli, or preexperimental to postexperimental change, on the evaluation or activity factors of the Semantic Differential. Significant "effects between stimuli" ($F = 6.48$, $p < .05$), and a significant interaction between stimuli and pre-post experimental change ($F = 11.68$, $p < .01$), were found in the analysis of the potency factor. Table 5 shows the mean potency ratings of the visual and auditory stimuli prior to, and following, isolation. The two kinds of stimuli were rated as equal in potency prior to isolation, but after isolation the visual stimuli were rated as less potent and the auditory stimuli as more potent.

AACL and SSCL Measures

There were no significant influences of GSR reactivity, prior experimental conditions, and no significant change in the AACL anxiety score from the preexperimental to the postexperimental testings.

There were no effects on the same "between-subjects" variables for the SSCL measure of somatic complaints; but the mean SSCL scores significantly increased from the pre- to postexperimental testings ($F = 31.54$, $p < .001$), going from a mean of 2.1, preisolation, to 4.5, postisolation.

DISCUSSION

The primary hypothesis, that persons who were most stressed during a previous isolation experience would respond more for visual and auditory sensory reinforcement in a second isolation experiment, was powerfully demonstrated. One explanation is that the prior GSR reaction was not specific to perceptual isolation but represented a general drive level. This general drive level might be measured by the subjects' levels of skin conductance just prior to isolation. There is some evidence that persons who are high in conductance during a prestress period are more reactive under stress than those who are initially low (Hord, Johnson, & Lubin, 1964). To test this possibility we performed the analysis contained in Table 3 again, only using the baseline conductance measure instead of the conductance change measure as the independent variable. The resulting F ratios were all small and nonsignificant. Therefore, it is the GSR reaction rather than the elevated prestress conductance levels which yielded the prediction of operant responding in the current experiment. Furthermore, the sensory reinforcement was a significant source of variation in the subject's responses. None of the subjects responded for mere kinesthetic self-stimulation, for example, with the switch in the middle position producing neither visual or auditory stimulation. Subjects responded significantly more for the visual reinforcement than for the auditory reinforcement and the difference between high and low GSR groups was more marked in their responding for visual reinforcement. This result is congruent with our result in the prior experiment with these subjects, where the group that was deprived of visual stimulation showed more nonspecific GSR reactions than the group deprived of auditory stimulation. The group deprived of both types of stimulation had more nonspecific GSRs than either of the partial isolation groups suggesting a summative effect of the two types of deprivation. Both experiments indicate the primacy of the visual sense in sensory deprivation for females (the visual versus auditory deprivation effects were not found for GSR reactions for males in another experiment [3]). Another indication of the greater role of visual deprivation than auditory deprivation is the fact that the responses for visual reinforcement increased during the third hour of deprivation while the responses for auditory reinforcement remained constant through the 3 hours of the experiment.

The increase in the responses for visual reinforcement during the third hour is interesting in view of the fact that the nonspecific GSR indication of stress in the prior isolation experiment did not differentiate the groups until the last hour and a half of the 3-hour period. It would appear that at least 3 hours of perceptual isolation are required to create a significant need for stimulation and the consequent stress reaction. Jones (1964) has demonstrated that the stimulus hunger for randomly varying series of stimuli continues to increase as deprivation time is extended.

---

[3] Unpublished manuscript, D. Vincent Biase, M. Zuckerman. Sex differences in responses to perceptual isolation.

The need for varied sensory stimulation appears to be an important source of the stress reaction to isolation. The fact that there are individual differences in this need suggests that there may be marked modification of the drive by learning. Just as some persons' "hunger" is increased more by learned "appetites," some persons' need for varied stimulation may be increased by inputs which change the levels of stimulation to which they become adapted. We need considerable study of the effects of these differences in optimal levels of stimulation in perceptual isolation and in other types of experiments. Although the experiment to experiment predictions proved to be a powerful method in this study, we should not neglect the attempt to develop other kinds of methods to assess this aspect of personality. Zuckerman, Kolin, Price, and Zoob (1964) have developed a Sensation-Seeking scale which may be useful in future research. The prediction of perceptual isolation reactions by the use of standard personality tests has not yielded impressive results (Hull & Zubek, 1962). Perhaps a "zeroing in" on the relevant aspects of behavior and personality, rather than the use of omnibus personality scales, will enable us to make more progress.

REFERENCES

BARON, A., & KISH, G. B. Low intensity auditory and visual stimuli as reinforcers for the mouse. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 1011–1015.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 95–98.

BUTLER, R. A. Incentive conditions which influence visual exploration. *Journal of Experimental Psychology*, 1954, **48**, 19–23.

BUTLER, R. A. The effect of deprivation of visual incentives on visual exploration motivation in monkeys. *Journal of Experimental Psychology*, 1957, **50**, 177–179.

BUTLER, R. A. The differential effect of visual and auditory incentives on the performance of monkeys. *American Journal of Psychology*, 1958, **71**, 591–593.

BUTLER, R. A., & ALEXANDER, H. M. Daily patterns of exploratory behavior in the monkey. *Journal of Experimental Psychology*, 1955, **48**, 247–249.

BUTLER, R. A., & HARLOW, H. F. Persistence of visual exploration in the monkey. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 258–263.

FOX, S. S. Self-maintained sensory input and sensory deprivation in monkeys: A behavioral and neuropharmacological study. *Journal of Comparative Physiological Psychology*, 1962, **55**, 438–444.

HARLOW, H. F., & ZIMMERMANN, R. R. Affectional responses in the infant monkey. *Science*, 1959, **130**, 421–432.

HORD, D. J., JOHNSON, L. C., & LUBIN, A. Differential effect of the law of initial value (LIV) on autonomic variables. *Psychophysiology*, 1964, **1**, 79–87.

HULL, J., & ZUBEK, J. P. Personality characteristics of successful and unsuccessful sensory isolation subjects. *Perceptual & Motor Skills*, 1962, **14**, 231–240.

JONES, A. Supplementary report: Information deprivation and irrelevant drive as determiners of an instrumental response. *Journal of Experimental Psychology*, 1961, **62**, 310–311.

JONES, A. Drive and incentive variables associated with the statistical properties of sequences of stimuli. *Journal of Experimental Psychology*, 1964, **67**, 423–431.

JONES, A., WILKINSON, J., & BRADEN, I. Information deprivation as a motivational variable. *Journal of Experimental Psychology*, 1961, **62**, 127–137.

KISH, G. B. Learning when onset of illumination is used as reinforcing stimulus. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 261–264.

LEVIN, H., & FORGAYS, D. G. Learning as a function of sensory stimulation of various intensities. *Journal of Comparative Physiological Psychology*, 1959, **52**, 195–201.

MOON, L. E., & LODAHL, R. M. The reinforcing effect of changes in illumination on lever pressing in the monkey. *American Journal of Psychology*, 1956, **69**, 288–290.

MYERS, A. K., & MILLER, N. E. Failure to find a learned drive based on hunger: Evidence for learning motivated by 'exploration.' *Journal of Comparative Physiological Psychology*, 1954, **47**, 428–436.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. New York: Oxford University Press, 1957.

PREMACK, D., COLLIER, G., & ROBERTS, C. L. Frequency of light-contingent bar pressing as a function of the amount of deprivation of light. *American Psychologist*, 1957, **12**, 411. (Abstract)

ROBINSON, J. Light as a reinforcer for bar pressing in rats as a function of adaptation, illumination level and direction of light change. *American Psychologist,* 1957, 12, 41–48.

VERNON, J. A., & MCGILL, T. E. Utilization of visual stimulation during sensory deprivation. *Perceptual & Motor Skills,* 1960, 11, 214. (Abstract)

ZUCKERMAN, M. The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology,* 1960, **24**, 457–462.

ZUCKERMAN, M. Perceptual isolation as a stress situation. *Archives of General Psychiatry,* 1964, **11**, 255–276.

ZUCKERMAN, M., ALBRIGHT, R. J., MARKS, G. S., & MILLER, G. L. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychological Monographs,* 1962, **76** (30, Whole No. 549).

ZUCKERMAN, M., & COHEN, N. Is suggestion the source of reported visual sensations in perceptual isolation? *Journal of Abnormal and Social Psychology,* 1964, **68**, 655–660. (a)

ZUCKERMAN, M., & COHEN, N. Sources of reports of visual and auditory sensations in perceptual isolation. *Psychological Bulletin,* 1964, **62**, 1–20. (b)

ZUCKERMAN, M., KOLIN, E. A., PRICE, L., ZOOB, I. Development of a sensation-seeking scale. *Journal of Consulting Psychology,* 1964, **28**, 477–482.

ZUCKERMAN, M., LEVINE, S., & BIASE, D. V. Stress response in total and partial perceptual isolation. *Psychosomatic Medicine,* 1964, **26**, 250–260.

(Received November 5, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 378–382

EXTINCTION OF A CONDITIONED MOTOR RESPONSE FOLLOWING HYPNOSIS[1]

JON M. PLAPP
Washington University

AND

WILLIAM E. EDMONSTON, JR.[2]
Washington University School of Medicine

The administration of a standard hypnosis induction procedure to 6 randomly assigned experimental Ss resulted in the disappearance of a recently acquired conditioned motor response. When hypnosis was terminated by another standard procedure the CR reappeared. On subsequent trials its appearance followed the usual extinction pattern. An equated control group (6 Ss) gave responses that followed the usual extinction pattern during both a "hypnosis-control" period and an identical 2nd block of extinction trials. Possible interpretations were considered and suggestions for further study made.

Noticing that the repetition of a conditioned stimulus (CS), without any presentation of the unconditioned stimulus (UCS), caused a lethargic or drowsy state in his animals, Pavlov (1927) developed a theory which clearly linked the conditioning process and hypnosis. He postulated that during conditioning the one area of the brain most directly affected by the CS became the center of an inhibitory process which spread to other parts of the brain, so long as the CS was presented monotonously and without reinforcement. The final stage of this process of inhibition was sleep, and one of the stages prior to sleep was what Pavlov termed hypnosis. Pavlov's concepts such as "cortical inhibition," "cortical irradiation," and cortical localization with respect to specific stimuli have been utilized as the predominant interpretive framework employed by subsequent Russian investigators in the hypnosis field. The generalized conclusions of certain of the Russian studies on hypnosis may not bear a close relation to the empirical findings; on the other hand, they may be considered in their own right, without reference to the theoretical framework.

There has been considerable disagreement over the effects of hypnosis on CRs. Russian workers have long viewed hypnosis as a valid field for experimental investigation and have devoted considerable attention to the measurement of physiological responses. Most Russian investigators have reported that a CR established during the waking state shows a diminution during hypnosis (Korotkin & Suslova, 1951, 1959, 1960; Levin, 1959; Pen & Jigarov, 1959; Povorinsky, 1959; Povorinsky & Traugott, 1959). That the CR may be restored upon termination of hypnosis, has been reported by Korotkin and Suslova (1951). Among Russian investigators reporting no effect of hypnosis on CR extinction has been Livshits (1959). In general, the reports of Russian experimentation have been characterized by a lack of attention to control groups, use of a very small number of subjects, and a minimal use of statistics. Procedures, both conditioning and hypnotic, often appear poorly standardized and are rarely reported in the detail which would permit adequate replications.

Many of these criticisms of the Russian work can, unfortunately, be applied with equal relevance to American studies. In contrast to the general Soviet findings of inhibition of CR under hypnosis, Americans have more frequently reported no change in the CR (McCrainie & Crasilneck, 1955; Scott, 1930). Other American investigators, reporting that CR changes were due to age regression (Le Cron, 1956) or to such postulated effects as hypnotic blindness (Ludholm, 1928) or deafness (Erickson, 1944) have clearly assumed that hypnosis per se has no effect on the CR.

No clear-cut conclusion as to the fate of a previously established CR following the induction of hypnosis seems justified from the above studies. Contradictory findings could be explained largely in terms of different methodologies and the use of different CRs. There is some evidence (Livshits, 1959; McCrainie & Crasilneck, 1955) that hypnosis may have an inhibiting effect on previously acquired vocabulary (or motor) CRs but not on involuntary (or perhaps more accurately, physiological) CRs. Finally, there has been insufficient use of controls, so that effects reported as attributable to hypnosis could be accounted for in terms of an increased state of relaxation in subjects, without introducing the more complex concept of hypnosis as an explanatory term (Orne, 1959).

---

[1] This research was supported in part by funds from National Institute of Health Grant MH 08677. The present paper was presented at the September 1964 meeting of the American Society of Clinical Hypnosis in Philadelphia, Pennsylvania.

[2] Now at Colgate University, Hamilton, New York.

The present study was designed to investigate one aspect of the problem: the effect of hypnosis on the extinction of a previously established motor CR. A control group was employed in an attempt to discover whether any obtained change in the CR could be attributed to hypnosis or simply to an increased state of relaxation. The experimental hypothesis was that during extinction trials following the establishment of a conditioned finger retraction response, subjects who had been hypnotized immediately prior to the start of the extinction period would produce CRs of significantly lesser magnitude and significantly greater latency than control subjects, who went through similar procedures but were simply asked to relax. It was further predicted that experimental subjects, when awakened and given a second set of extinction trials, would give CRs which did not differ significantly from those produced by control subjects in a parallel second extinction period.

METHOD

Subjects

From approximately 30 Washington University and Medical School volunteer students, 12 subjects were chosen and equally divided into experimental (hypnosis) and control groups. Fourteen of the original 30 were excluded because they were unable to meet the conditioning criterion; the others failed to keep their appointments.

Assignment of each subject to an experimental or control group was made on a random basis. Each group contained four females and two males, and the two groups were equated with respect to mean age (19) and years of education (14).

Apparatus

The conditioning apparatus consisted of (*a*) an Eico Model 377 Audio Generator, adjusted to deliver the CS (a 2000 cycle per second, 80 decibel sine wave) to an amplifier which was placed approximately 3 feet behind the subject, (*b*) a Burdick muscle stimulator (Model MS-2) which delivered an UCS of up to 60 miliamperes. The UCS was delivered to the subject through two silver electrodes attached to the middle finger of the subject's right hand. Finger movements resulted in variations in the light reaching selenium cells, and activated one channel of an Offner polygraph, resulting in pen deflections which bore a linear correspondence to the finger movements. All fingers on the subject's right hand except for his middle finger were strapped down firmly and another strap was placed around the subject's forearm. Marker pens recorded onset, duration, and cessation of CS and UCS. The corresponding time periods, as well as the duration of the interval between CS and UCS, were controlled by automatic Hunter timers. The conditioning apparatus was operated from a one-way mirrored room next to the one in which the subject was seated. A red spot (0.5-inch diameter), placed on the wall in front of the subject's chair, approximately 3 feet above his eye level, served as the "target" in the hypnosis induction procedure.

Procedure

Both groups received identical conditioning procedures. After each subject was seated in a comfortable chair electrodes were attached, a large base electrode at the wrist, and the shock electrode on the middle finger of the right hand. The thread which led to the finger retraction recording apparatus was attached to the subject's middle finger.

Next, conditioning instructions, modeled after those of Lindley and Moyer (1961) were read. No subject responded with finger retraction to the initial presentations of tone. The shock was then adjusted to the strongest level of intensity the subject could tolerate. Learning trials were then presented, one in every block of five trials being a CS-alone presentation (test trial). The CS was presented for .5 second, followed by a .5-second interval, followed in turn by the UCS which lasted for .2 second. If it appeared that the subject was adapting to the shock level its intensity was increased. Intertrial intervals were randomly scheduled between 20–40 seconds. A CR was defined as a pen deflection of no less than 1 millimeter from the baseline, occurring within 1 second after the onset of the CS. The criterion of conditioning was that the subject should give four conditioned responses in five consecutive test trials.

After conditioning, experimental subjects were given the following sections of the Stanford Hypnotic Susceptibility Scale (SHSS), Form A (Weitzenhoffer & Hilgard, 1959): (*a*) eye closure, (*b*) hand lowering (left hand), (*c*) arm rigidity (left), (*d*) eye catalepsy. The standard instructions were read to the subject in a quiet, rhythmic voice, the experimenter being seated behind and to the left of the subject, out of his line of vision. On each of the items *a–d* the subject's performance was scored one or zero. If the subject obtained a total score of zero he was not considered hypnotized; any score from one to four was accepted as indicating hypnosis. No experimental subject obtained a total hypnotic susceptibility score of zero; one obtained a score of one; one, three; and the three remaining subjects, scores of four.

Following the above induction procedures, hypnosis group subjects were given the following instructions:

> Remain deeply relaxed and pay close attention to what I am going to tell you. In a moment I will be leaving the room for a while. All the time I am gone you will remain just as deeply relaxed as you are now. You will not wake up until I return and give you the instructions to awake. Remember, all the time I am gone you will continue to remain in a very deep state of relaxation. You will hear me leave, and you may hear me in the next room, but you will continue to feel deeply, comfortably relaxed, and you will not awake until I return and give you the instructions to awake. Now I am going to leave.

Twelve extinction trials were then administered to the hypnotized subject, an extinction trial being defined as the presentation of a CS without the UCS. Next, the arousal instructions from the SHSS were read. After the subject had been awakened another 12 extinction trials were administered.

Control and experimental groups received the same conditioning and extinction procedures, but instead of the hypnosis induction procedures ("eye closure" section of the SHSS), controls received the following instructions after conditioning:

> Now we come to the second part of the experiment. You are *not* going to be hypnotized in this experiment. Instead, I am simply going to ask you to relax, get as comfortable as you can, and go through certain simple procedures as I tell you to. I just want you to quite consciously follow the instructions I am going to give you.

Controls were instructed to make themselves comfortable and close their eyes. After a period of 20 minutes they were told to perform voluntarily the behaviors described above for hypnosis subjects (beginning with item *b* of the SHSS) and then the first 12 extinction trials were administered. Following this, control subjects were instructed to open their eyes and sit up again, and finally the second 12 extinction trials were presented.

RESULTS

In general, the records of control and experimental subjects demonstrated that subjects who had been hypnotized immediately prior to the start of extinction produced significantly fewer CRs during this period than control subjects, and that when hypnosis was terminated experimental and control subjects did not differ significantly (see Figure 1). To test for initial differences between the two groups in number of trials to criterion, a Mann-Whitney U test was applied. No significant difference was obtained (M of 38.8 and 36.0). The extinction records of each subject were scored independently by the two experimenters with a high degree of correspondence between their results (Pearson r = .95). Figure 1 presents the number of subjects giving CRs on each extinction trial (for the two extinction periods). The maximum number of CRs possible for either group on any trial was six, one for each subject. The difference between experimental and control subjects during the first extinction period is clear. Only a single hypnosis subject responded at all to CS presentation during the first extinction period, while either four or five control subjects gave CRs on each trial. The sole hypnosis subject who responded was in fact

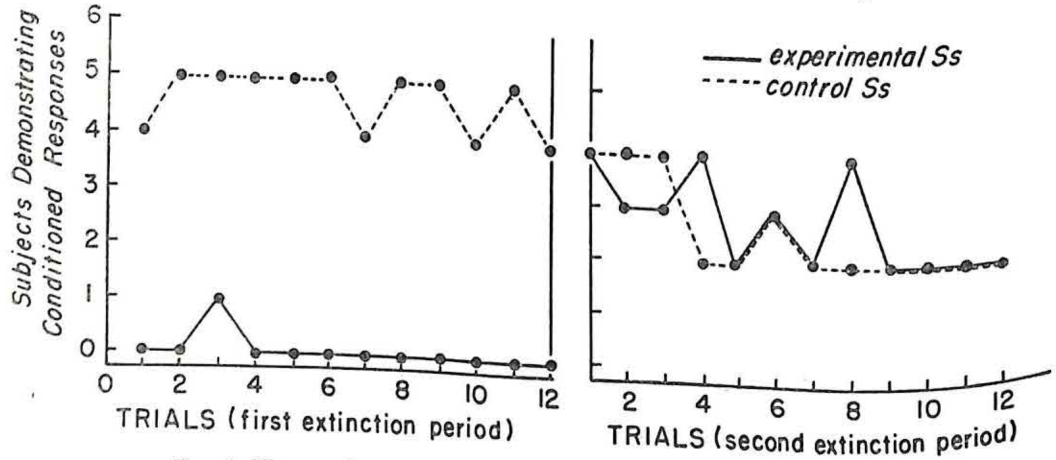


FIG. 1. The number of subjects giving CRs on each extinction trial.

that subject who obtained the lowest hypnotic susceptibility score, a score of one. During the second extinction period, after hypnosis subjects had been awakened, there was a reappearance of the CR for four experimental subjects on the initial trial and a subsequent gradual reduction in the number of CRs given. The groups were found not to differ significantly in number of CRs evoked during this second extinction period. A single experimental subject failed to respond during the second extinction period, and also one control subject failed to respond during either the first or second blocks of extinction trials. Since the CR was almost completely absent during the first extinction period for the experimental subjects, examination of the data in terms of CR amplitude and latency as had been originally proposed, was unnecessary.

DISCUSSION

If, as has been shown in this experiment, hypnosis does have an effect on the CR, it would seem that the type of response conditioned may be an important determinant of the effect obtained. The findings of a study by McCrainie and Crasilneck (1955) suggest such an inference. They found that all 12 subjects responded to the CS under hypnosis as they had in the normal state. Following age regression, the CR disappeared in the 6 subjects in whom a voluntary CR had been established, but failed to be influenced in the case of the 6 subjects in whom an involuntary response had been conditioned. Also relevant to the question of whether hypnosis has any effect on an established involuntary CR was the investigation by Livshits (1959), who conditioned differential vascular dilation and contraction responses to a buzzer and a bell before inducing hypnosis. Although he began with an impressive number of subjects (43), failures in conditioning and/or hypnosis reduced the *N* considerably (at least to 19), and the author presented his results mainly in intraindividual terms. Apparently there was much variation in the conditioning, and particularly in the hypnosis induction procedures used. Despite the general conclusion that "conditioned reflexes elaborated in the waking state were preserved in the state of hypnosis [p. 752]," clear evidence of this was presented only for a single subject.

Should a voluntary, but not an involuntary CR be inhibited during hypnosis, there is a strong possibility that this inhibition is part of an effect of hypnosis on the motor responsiveness of the subject, rather than being an interference with the ability to perceive the CS. And if, as seems to have been the case, most or all of the CRs which were reported to have disappeared or shown a reduction in amplitude under hypnosis were motor responses, it could be argued that the state of relaxation which acompanies hypnosis interferes with the generalized responsiveness of the subject. An adequate test of this possibility would necessitate the use of a control group who, though not hypnotized, were "relaxed" to the same extent as the hypnotized subjects. Investigation of the effects of employing such a control condition was one of the aims of the present study.

Since control subjects were presumably as relaxed as experimental subjects during first extinction period (they had sat with eyes closed for 20 minutes), the differences between the CRs produced by the two groups can be attributed to the specific hypnosis induction instructions received by the experimental group. Possibly, it was the repetitive, monotonous auditory stimulation from the experimenter which produced the differential effect on the CR, as would be suggested by the work of Dittborn and Kline (1958). Or it may have been the initial statement by the experimenter to experimental subjects that they were to be hypnotized and to control subjects that they were not to be, which led to the present results. The latter possibility would be consistent with the findings of Barber and Calverley (1964b) that subjects told they were in the hypnosis group were more responsive to suggestions than others told they were in the control group. Whether explanation of the present results is to be sought in terms of habituation resulting from repetitive stimulation or in the expectations of subjects, there is evidence here that the hypnotic state is not simply a point on a continuum from "alert wakefulness" through "relaxation" to "sleep," as Pavlov and other Russian investigators have suggested. The present results also make necessary a reconsideration of certain conclusions regarding the effects of specific suggestions made to subjects during hypnosis. If hypnosis alone can result in the disappearance of a CR, this may help to explain some of the results of studies of "hypnotic blindness," "hypnotic deafness," and "age regression" (Erickson, 1944; Le Cron, 1956; Lundholm, 1928). The need, in studying the effects of such hypnotic suggestions, for control groups of subjects who are simply hypnotized is thus underlined by the results of the present experiment.

Some questions may be raised about the adequacy of the procedures used in the present investigation. The treatment of control subjects during the interval which was intended to approximate that of the hypnosis period for experimental subjects, may not have served its purpose adequately. Although it was attempted to match the two groups, control subjects differed in terms of receiving shorter, less repetitive introductory instructions prior to performing the same behaviors as experimental subjects ("eye-closure" from the SHSS was omitted for controls). This fact makes it impossible to specify with complete assurance whether the behavior of experimental subjects during the first extinction period was due to "hypnosis" or simply to the longer and more detailed instructions they had received. In addition, the fact that the situation was defined to experimental (but not control) subjects as hypnosis, together with possible differential variations in experimenter's tone of voice, may have implied to experimental (but not to control) subjects that they were not expected to show the CR during the first extinction period. There is a need for further studies (e.g., Barber & Calverley, 1964a) which will vary systematically the instructions given to hypnosis and nonhypnosis groups, and also for investigations of the effects of hypnosis on voluntary as well as involuntary CRs. While certain questions may be raised about the meaning of the present findings, there seems to be no doubt that extinction of a motor CR was clearly affected by hypnosis.

REFERENCES

BARBER, T. X., & CALVERLEY, D. S. Effect of E's tone of voice on "hypnotic-like" suggestibility. *Psychological Reports,* 1964, **15**, 139–144. (a)

BARBER, T. X., & CALVERLEY, D. S. Toward a theory of hypnotic behavior: Effects on suggestibility of defining the situation as hypnosis and defining response to suggestions as easy. *Journal of Abnormal and Social Psychology,* 1964, 68, 585–592. (b)

DITTBORN, J. M., & KLINE, M. V. An instrument for the measurement of sleep induction. *Journal of Psychology,* 1958, **46**, 277–278.

ERICKSON, M. H. A study of clinical and experimental findings on hypnotic deafness. II Experimental findings with a conditioned response technique. *Journal of General Psychology,* 1944, 31, 191–212.

KOROTKIN, I. I., & SUSLOVA, M. M. *Zhurn Vyssh Nervn. Deyatel' Nosti,* 1951, Vol. I (4).

KOROTKIN, I. I., & SUSLOVA, M. M. Changes in conditioned and unconditioned reflexes during suggestion states in hypnosis. In, *The Central Nervous System and Human Behavior.* Princeton, N.J.: Public Health Service, U.S. Department of Health, Education, and Welfare, 1959.

KOROTKIN, I. I., & SUSLOVA, M. M. Comparative effects of suggestion in the waking state and in hypnosis. *Pavlov Journal of Higher Nervous Activity,* 1960, **10**, 185–192.

LE CRON, L. M. A study of age regression under hypnosis. In L. M. Le Cron (Ed.), *Experimental hypnosis.* New York: Macmillan, 1956.

LEVIN, S. In K. Platonov (Ed.), *The word as a physiological and therapeutic factor.* Moscow: Foreign Languages Publishing House, 1959. p. 75.

LINDLEY, R. H., & MOYER, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *Journal of Experimental Psychology,* 1961, **61**, 82–88.

LIVSHITS, L. S. The investigation of the higher nervous activity of man in hypnosis in relation to chronic alcoholism. *Pavlov Journal of Higher Nervous Activity,* 1959, **9**, 745–753.

LUNDHOLM, H. An experimental study of functional anesthesias in hypnosis. *Journal of Abnormal and Social Psychology,* 1928, **23**, 337–355.

MCCRAINIE, E. J., & CRASILNECK, H. B. The conditioned reflex in hypnotic age regression. *Journal of Clinical and Experimental Psychopathology,* 1955, **16**, 120–123.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology,* 1959, **58**, 277–299.

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Universities Press, 1927.

PEN, R., & JIGAROV, M. In K. Platonov (Ed.), *The word as a physiological and therapeutic factor.* Moscow: Foreign Languages Publishing House, 1959. p. 75.

POVORINSKY, Y. (1937). In K. Platonov (Ed.), *The word as a physiological and therapeutic factor.* Moscow: Foreign Languages Publishing House, 1959. p. 75

POVORINSKY, Y., & TRAUGOTT, N. In K. Platonov (Ed.), *The word as a physiological and therapeutic factor.* Moscow: Foreign Languages Publishing House, 1959. p. 75.

SCOTT, H. D. Hypnosis and the conditioned reflex. *Journal of General Psychology,* 1930, **4**, 113–129.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

(Received November 9, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 383–389

TRANCE-SUSCEPTIBILITY, INDUCTION-SUSCEPTIBILITY, AND ACQUIESCENCE AS FACTORS IN HYPNOTIC PERFORMANCE[1]

EVELYN LEE-TENG

California Institute of Technology

Scores from 103 Ss on a 60-item inventory of attitudes and experiences outside of hypnosis, called the Hypnotic Characteristics Inventory (HCI), along with sum-true scores on an abbreviated form of the MMPI, and interviewer predictions, were entered into a correlational matrix together with later-obtained hypnotic susceptibility scores derived from SHSS Forms A and C. The matrix was factor-analyzed and rotated to a hypothetical structure. Hypnotic susceptibility was characterized by 2 oblique factors, 1 termed trance-susceptibility, the other induction-susceptibility. Questionnaire measures, which were found compounded with an acquiescence tendency, were correlated with trance-susceptibility but not induction-susceptibility, indicating a limitation in hypnotic prediction from questionnaires.

Attempts to relate personality characteristics to hypnotizability have met with very little success (e.g., Barber & Glass, 1962; Deckert & West, 1963). One rather promising approach, introduced by Shor (1960), made use of a Personal Experience Questionnaire (PEQ), in which the subjects reported experiences in everyday life that resembled those called for within hypnosis. Shor, Orne, and O'Connell (1962) extended the original 44-item PEQ to a 149-item version, and found 45 items which on the basis of their "intensity scores" correlated with hypnosis scores significantly at the .05 level. When these items were cross-validated on a new sample of 25 subjects, the sum of the intensity scores correlated .44 with scores in hypnotic susceptibility. Taking his lead from Shor's work, and using Shor's instructions and 18 of his items, Ås constructed a 60-item scale to tap the types of experiences which he felt to be related to hypnosis. He and his collaborators studied the internal structure of the Experience Inventory (EI), and related it to hypnosis (Ås, 1962; Ås & Lauer, 1962; Ås, O'Hara, & Munger, 1962). Their scoring differed from that of Shor and his collaborators in that a simple frequency count of "yes," "no," and "?" answers were correlated with hypnosis, counting one for a "yes" answer and one half for a "?" reply. In a sample of 100 male students Ås found a correlation of .37 with the Stanford Hypnotic Susceptibility Scale, Form C (SHSS-C), a scale standardized by Weitzenhoffer and Hilgard (1962); the correlation within a sample of 102 women students turned out to be .33 (Ås, 1962).

A puzzling problem arose when Ås attempted to make use of an item analysis in order to improve the predictive value of his EI. While a shortened scale could be found that would yield a much higher correlation on the sample on which the item selection was based, the new scale, on replication, was no more predictive than the scale composed of the discarded items (Ås, 1962). This suggests that some content-free, stylistic aspects of the inventory may be influential in determining the correlations that are found.

The present study was undertaken with the following objectives: (*a*) to clarify the aspects of personality and experience measured by an inventory of the Shor-Ås type through grouping all items into a small number of distinguishable subscales, adding items as necessary to balance the number in each subscale; and (*b*) to determine the aspects of pencil-and-paper inventories and responses within hypnotic suggestibility scales that limit the intercorrelations to be expected. For these purposes a modified inventory was constructed and the results were intercorrelated both with hypnosis and with other measures in order to ascertain more clearly the factors involved.

---

[1] This report is rewritten from a doctoral dissertation prepared at Stanford University under the general direction of Ernest R. Hilgard and carried out in the Laboratory of Hypnosis Research (Lee, 1963). The author wishes to acknowledge the aid of Santo F. Camilleri, Perry London, and Quinn McNemar, who served with E. R. Hilgard as dissertation consultants. The author wishes to thank E. R. Hilgard also for his aid in the revision of the dissertation for publication. The research reported is part of the ongoing program of research on hypnosis supported by Grant Number M-3859 to E. R. Hilgard from the National Institute of Mental Health.

HYPNOTIC CHARACTERISTICS INVENTORY

The newly constructed inventory, which leans heavily upon those of Shor and Ås, will be referred to as the Hypnotic Characteristics Inventory (HCI) because it is incompletely described as an experience inventory. On the basis of review of data earlier obtained within the Laboratory of Hypnosis Research, including coded interviews as well as inventory results, five dimensions were selected as being promising: conformity versus autonomy, trancelike experiences, role-playing, impulsivity versus rationality, and concentration-absorption. The terms in most cases suggest the type of response expected to be related to hypnosis; in the case of the two scales which represent continua, the first member of the pair of terms was conjectured to characterize the hypnotizable person (i.e., conformity and impulsivity). In order to produce five subscales each with 12 items, 30 newly constructed items were added to 30 selected from the Shor-Ås inventories. In order to check on the face-validity of item placement, 10 staff members and assistants were asked to sort the 60 items into five piles, each representing the most appropriate subscale. The average agreement was 79% with respect to the placement of an item into its appropriate category. While there is, therefore, a little ambiguity, on the whole the distinctions among the categories appeared satisfactory.

The following items are illustrative of those composing the various subscales:

1. Conformity versus Autonomy—"In general, do you prefer to follow rather than to direct?" (Yes)

2. Trancelike experiences—"When you dance, do you often feel that the music and mood are being expressed through your movements, while you yourself fade into the background?" (Yes)

3. Role-playing—"While watching a movie or show do you sometimes become so involved that you feel yourself participating in the action?" (Yes)

4. Impulsivity versus Rationality—"When faced with a decision, do you usually ponder and weigh all aspects carefully?" (No)

5. Concentration and Absorption—"Do you often have trouble in keeping your mind on what you should do?" (No)

The subjects used in this study were men and women students in the introductory psychology classes at Stanford University. The inventory measures were obtained early in the term prior to the start of the hypnosis experiments. First, the retest reliability was obtained from a sample of 47 subjects who took the test twice, the second time 2 weeks after the first. The retest reliabilities of the five subscales were: .73, .84, .74, .75, and .73, respectively, and for the 60-item inventory, .83. Secondly, the whole inventory was given to a sample of 236 subjects, from whom the 103 hypnotic subjects were later to be enlisted. Biserial correlations between individual items and the subscale to which it belonged yielded positive correlations throughout, all significant at the .05 level or higher. With this degree of internal consistency and the determined reliabilities of the subscales, the inventory appeared to be ready for use.

It was important to establish, first, that the previously found relationships with hypnosis could be replicated. As shown in Table 1, each of the HCI subscales correlates with the total HCI score (minus that subscale), and for three of the five subscales the correlation

TABLE 1

CORRELATION BETWEEN THE HCI SUBSCALE SCORES TOTAL HCI, AND HYPNOSIS, FORM C

Subscale	Number of items	HCI (with subscale omitted)	SHSS Form C
1. Conformity	12	20*	04
2. Trancelike	12	45**	19*
3. Role-playing	12	59**	38**
4. Impulsivity	12	36**	26**
5. Concentration	12	29**	12
Total HCI	60		33**

Note.—$N = 103$.
* $p < .05$.
** $p < .01$.

with hypnotic susceptibility is significant; the HCI as a whole correlates significantly with hypnosis.

FACTORIAL ANALYSIS OF THE HCI SUBSCALES IN RELATION TO OTHER MEASURES

A number of measures were obtained to be used in a factor analysis designed to answer the questions with which the study began. It should be pointed out that the purpose of such an analysis is not that of producing an efficient measure by which to predict hypnotic susceptibility. Anyone familiar with hypnosis knows that the most efficient predictive measure is a short work sample of hypnosis itself, a 10-minute practice session yielding higher predictions than anyone has been able to produce to date by personality measures or interviews. There are, however, things to be learned from the modest correlations that are obtained, and what is learned may lead to a better understanding of hypnotic phenomena even if not very effective in selecting susceptible candidates for hypnosis.

With such analytical purposes in mind, a factor analysis was undertaken of scores on 10 variables, each of which yielded a distribution permitting product-moment correlations with other variables. The 10 variables were as follows, with the possible scores on each:

HCI Subscales (Scores 0–12 on each)

1. HCI-Conformity
2. HCI-Trancelike Experiences
3. HCI-Role-Playing
4. HCI-Impulsivity
5. HCI-Concentration-Absorption

Other Personality Measures

6. Sum-True Scores on the MMPI. The score consisted in the number of items endorsed "true" on a 235-item abbreviated MMPI, thus yielding a "yea-saying" or "acquiescence" score. The possible range is, of course, 0–235.

7. Interviewer Predictions. A relatively unstructured interview in advance of hypnosis, covering many aspects similar to those sampled in the HCI (plus many other topics), led to predictions of hypnotic susceptibility in advance of hypnosis. These were made on the scale of 0–12, which was the score range of the whole SHSS-C. The interviewers had no knowledge of the score on the HCI, and the hypnotists did not know the interviewer predictions.[2]

Hypnotic Susceptibility Subscales (Scores 0–6)

8. SHSS-A, unique items. These are scores on the 6 of the 12 items from Form A (Weitzenhoffer & Hilgard, 1959) which do not appear in Form C. They are from a first day of hypnosis, and are mostly motor items.

9. SHSS-C, unique items. These are scores on the 6 of the 12 items from Form C that do not appear on Form A. They are from a second period of hypnosis. The items are largely of cognitive type (dream, regression, hallucinations).

10. SHSS-C, items common to Forms A and C. While the scores are all taken from Form C, they are for items which occur also on Form A, and fall in content somewhat between the other sets of items, in that they include some motor items, but also amnesia and hallucination.

The reasons for including the various items are fairly obvious, but some details need to be filled in. The subscales of the HCI are there to tell us something about the uses and limitations of such an inventory; the regroupings of the hypnotic susceptibility items should throw some light on a frequently found higher correlation of predictive measures with Form C than with Form A (Weitzenhoffer & Hilgard, 1962); the MMPI measure, as an acquiescence measure, should tell the extent to which such a stylistic tendency enters into HCI results, and the interviews should yield related measures, but less contaminated by the verbal styles of the subjects in responding to pencil-and-paper tests calling for yes-no answers.

The intercorrelations of the measures are shown in Table 2. Values in the diagonal are

---

[2] The interviewer predictions were furnished by Josephine R. Hilgard, in charge of the interview project. The details of the interview results have not yet been published, but some of the hypotheses being tested and some case material appear in Hilgard and Hilgard (1962), and some preliminary analyses in Hilgard (1964, 1965).

TABLE 2

CORRELATION MATRIX WITH SMCS IN THE DIAGONAL

Variable	1	2	3	4	5	6	7	8	9	10
1. HCI: Conformity	$\boxed{07}$	10	21	09	16	07	17	−08	05	04
2. HCI: Trancelike		$\boxed{27}$	49	32	19	32	11	13	20	15
3. HCI: Role-playing			$\boxed{51}$	41	32	50	26	20	36	35
4. HCI: Impulsivity				$\boxed{23}$	09	35	21	21	25	24
5. HCI: Concentration					$\boxed{20}$	−05	12	00	10	13
6. MMPI: Sum-true						$\boxed{35}$	06	10	28	20
7. Interviewer prediction							$\boxed{19}$	28	24	19
8. Form A: unique								$\boxed{65}$	63	76
9. Form C: unique									$\boxed{58}$	73
10. Form C: common										$\boxed{71}$

Note.—$N = 103$.

the squared multiple correlations (SMC), of each of the variables with the remaining 9 variables. The SMC is a lower bound estimation of communality, h^2, and is often recommended as an approximation to it (Harmon, 1960, pp. 89–91).

While a number of inferences can be drawn from Table 2, these will become clearer after the results of factor analysis are presented. One point that will be lost in the factor analysis is the indication of which items of the SHSS are responsible for the usual higher correlations between personality measures and SHSS-C over those found with SHSS-A (Weitzenhoffer & Hilgard, 1962). Examining the correlations between the unique items of Form A and those of Form C (Columns 8 and 9) with the scores from the various inventories (Items 1 to 6) it will be noted that *in every case* the correlation is higher with the unique items of Form C (Column 9) than with the unique items of Form A (Column 8). It thus appears that the unique items of Form C (dream, regression, hallucinations) are the ones producing the higher correlations with pencil-and-paper tests, by contrast with the predominantly motor tasks reflected in the unique items of Form A. The fact that most of the correlations with the common items (Column 10) fall nearer to the unique Form C correlations (Column 9) than to the unique Form A ones (Column 8) is not in contradiction, in that these common items include such cognitive ones as amnesia and a fly hallucination.

Table 3 gives the results of factor analysis, prior to rotation. Loadings on the first three factors are presented; the percentage of variance explained is 43%. When factorization was carried to the tenth "factor" the variance explained was but 45%; the factor loadings beyond the third factor are negligibly small and are therefore not presented. The preliminary factor solution presented in Table 3 was derived solely from the mathematical criterion of ellipsoidal fit; such an orthogonal axis system does not mean much psychologically, and hence rotations are undertaken.

ROTATION TO A HYPOTHETICAL FACTOR STRUCTURE

A number of possibilities exist for attempting to derive psychological meaning through rotating the axes, once a table such as

TABLE 3

FACTOR ANALYSIS BY THE PRINCIPAL COMPONENTS METHOD USING SMC AS THE ESTIMATION OF h^2

Variable	Factor loading			h^2
	I	II	III	
1. HCI: Conformity	13	23	23	12
2. HCI: Trancelike	41	39	−04	32
3. HCI: Role-playing	64	47	04	63
4. HCI: Impulsivity	45	26	−10	28
5. HCI: Concentration	21	22	39	24
6. MMPI: Sum-true	44	35	−36	45
7. Interviewer prediction	35	04	23	18
8. Form A: unique	69	−48	−01	71
9. Form C: unique	75	−25	−02	63
10. Form C: common	79	−38	−04	77
Percentage variance	28	11	4	
Cumulative percentage variance	28	39	43	43

Note.—$N = 103$.

Table 3 has been arrived at. In the original dissertation (Lee, 1963) an orthogonal rotation was undertaken, yielding two chief factors: one reflecting the hypnotic susceptibility variables, the other the pencil-and-paper inventories, which are rather similar to the results found from another sample by Ås and Lauer (1962). Most of the measures, however, fell in the same quadrant when plotted in the plane of the two factors so that it became apparent that simple structure was lacking, and that a rotation which would permit oblique axes would yield additional pertinent information. Such a rotation was also performed, and its results will be presented here.

The method followed is that of Horst (1956). Factors are hypothesized in advance, and the results are then examined for goodness of fit of the obtained rotated factor matrix.

The procedure is as follows. Instead of the empirically obtained values, the numeral "1" is placed in the matrix wherever a factor loading is hypothesized, "0" wherever a loading on that factor is not expected. This has been done for three hypothesized factors in Table 4. The first hypothesized factor is named "trance-susceptibility," representing a subject's capability to experience the trance states. This factor is expected to enter all of the 10 variables with the exception of HCI-Conformity. The second factor, "induction-susceptibility," represents a subject's susceptibility to the induction of hypnosis. This second factor is expected to load on HCI-Conformity, the interviewer predictions, as well as the three hypnotic measures derived from SHSS Forms A and C. Finally, a third factor, named "acquiescence," represents a subject's tendency to endorse the "yes" or "true" answer option on the pencil-and-paper inventories, and is expected to load on scores from the five HCI subscales and the MMPI sum-true score. The rotational procedure transforms the original axis system of the factor analysis to one which best fits the axes of the hypothetical matrix. There are no restrictions to orthogonality. The result is an intermediate matrix, with somewhat inflated factor loadings because of the values of 1 and 0 used in the hypothetical matrix. Hence this matrix is then transformed to yield a matrix with h^2 values equal to those of the original matrix. The results of these rotations and transformations are given in Table 5.

Considering only the transformed matrix, it can be seen how well the hypothesized factors emerge. The conjecture that all but the first variable would load on the first factor is supported by factor loadings of .22 to .49 for the nine variables. Thus trance-susceptibility is widely characteristic of what the inventories and hypnosis have in common. Induction-susceptibility, the second factor, loads .41 to .56 for the three SHSS variables,

TABLE 4

HYPOTHETICAL FACTOR STRUCTURE (H)

Variable	Factor		
	I Trance-Suscepti-bility	II Induction-Suscepti-bility	III Acquies-cence
1. HCI: Conformity	0	1	1
2. HCI: Trancelike	1	0	1
3. HCI: Role-playing	1	0	1
4. HCI: Impulsivity	1	0	1
5. HCI: Concentration	1	0	1
6. MMPI: Sum-true	1	0	1
7. Interviewer prediction	1	1	0
8. Form A: unique	1	1	0
9. Form C: unique	1	1	0
10. Form C: common	1	1	0

TABLE 5

FACTOR STRUCTURE AFTER ROTATING TO THE HYPOTHETICAL MODEL

Variable	Intermediate matrix				Transformed matrix			
	I	II	III	k	I	II	III	h²
1. HCI: Conformity	24	12	30	64	15	08[a]	19[a]	12
2. HCI: Trancelike	49	03	51	61	30[a]	02	31[a]	32
3. HCI: Role-playing	74	18	69	57	42[a]	10	39[a]	63
4. HCI: Impulsivity	48	07	40	63	30[a]	04	25[a]	28
5. HCI: Concentration	35	28	36	63	22[a]	18	23[a]	24
6. MMPI: Sum-true	43	−15	41	90	39[a]	−14	37[a]	45
7. Interviewer prediction	38	34	22	57	22[a]	19[a]	13	18
8. Form A: unique	48	61	−15	92	44[a]	56[a]	−14	71
9. Form C: unique	61	53	08	77	47[a]	41[a]	06	63
10. Form C: common	62	65	−02	79	49[a]	51[a]	−02	77

[a] Hypothesized to load on the factor.

and .19 for the interviewer predictions, but fails to load (.08) as predicted on HCI-Conformity. In fact, none of the pencil-and-paper inventories load on this factor. Finally, the acquiescence factor is represented by loadings of .19 to .39 on all the inventories, including the subscales of the HCI and MMPI sum-true scores.

The computer program which carries out the indicated rotation also yields the intercorrelations of the factors, as shown in Table 6. These intercorrelations well summarize the major findings of this analysis by showing that trance-susceptibility is correlated with both of the other factors (.54 and .70), and accounts for the success with which experience inventories predict hypnosis. These inventories, however, are loaded with an acquiescence factor (Factor III) which fails to correlate with the second factor, induction-susceptibility. Thus it is clear why improving these inventories is always likely to miss something important in hypnosis.

DISCUSSION

The original hope to specify the content areas of the HCI responsible for the predictive significance of experience inventories and hypnotic characteristics tests has been dampened by the prominence of the acquiescence tendency, which makes it difficult to distinguish between content and style. Hence it is preferable at this stage to consider the HCI as a whole, as representing the inventory approach to hypnotic susceptibility.

The factor analysis permitting oblique rotation resulted in the breaking up of hypnotic performance as represented in the Stanford Hypnotic Susceptibility Scales into two parts: trance-susceptibility and induction-susceptibility. Because these two factors are oblique and correlated to the extent of .54, orthogonal rotations may fail to separate them. When inventory items are included in the correlational matrix, it turns out that these items are more highly correlated with the trance-susceptibility component (.22 to .42) than with the induction-susceptibility component (−.14 to .18). The yea-saying tendency (acquiescence) represented by the inventories appears to be the factor responsible for this, for as a factor it correlates .70 with the trance-susceptibility factor and −.02 with the induction-susceptibility factor. What is here called the acquiescence tendency is not entirely content-free, because the inventory items selected do bear some face similarity to hypnosis, and the more "yes" answers, the more likely the subject has experiences outside hypnosis which mimic those called for during hypnosis. Yet the stylistic tendency is so strong that, within a pool of relevant items, it is impossible at present to indicate

TABLE 6

INTERCORRELATIONS BETWEEN THE ROTATED FACTORS

Factor	I	II	III
I. Trance susceptibility		54	70
II. Induction susceptibility			−02
III. Acquiescence			

more specifically just what kinds of items are most pertinent for the prediction of hypnosis. The distinction between trance-susceptibility and induction-susceptibility provides a new direction for research on the personality correlates of hypnosis.

REFERENCES

ÅS, A. Non-hypnotic experiences related to hypnotizability in male and female college students. *Scandinavian Journal of Psychology,* 1962, **3,** 112–121.

ÅS, A., & LAUER, LILLIAN W. A factor analytic study of hypnotizability and related personal experiences. *International Journal of Clinical and Experimental Hypnosis,* 1962, **10,** 169–181.

ÅS, A., O'HARA, J. W., & MUNGER, M. P. The measurement of subjective experiences presumably related to hypnotic susceptibility. *Scandinaian Journal of Psychology,* 1962, **3,** 47–64.

BARBER, T. X., & GLASS, L. B. Significant factors in hypnotic behavior. *Journal of Abnormal and Social Psychology,* 1962, **64,** 222–228.

DECKERT, G. H., & WEST, L. J. The problem of hypnotizability: A review. *International Journal of Clinical and Experimental Hypnosis,* 1963, **11,** 205–235.

HARMON, H. H. *Modern factor analysis.* Chicago: Univer. Chicago Press, 1960.

HILGARD, E. R. The motivational relevance of hypnosis. In *Nebraska symposium on motivation: 1964.* Lincoln: Univer. Nebraska Press, 1964. Pp. 1–46.

HILGARD, E. R. *Hypnotic susceptibility.* New York: Harcourt, Brace & World, 1965.

HILGARD, JOSEPHINE R., & HILGARD, E. R. Developmental-interactive aspects of hypnosis: Some illustrative cases. *Genetic Psychology Monographs,* 1962, **66,** 143–178.

HORST, P. A simple method of rotating a centroid factor matrix to a simple structure hypothesis. *Journal of Experimental Education,* 1956, **24,** 251–258.

LEE, EVELYN M. A questionnaire measure of hypnotic characteristics and their relationship to hypnotizability. Unpublished doctoral dissertation, Stanford University, 1963.

SHOR, R. E. The frequency of naturally occurring "hypnotic-like" experiences in the normal college population. *International Journal of Clinical and Experimental Hypnosis,* 1960, **8,** 151–163.

SHOR, R. E., ORNE, M. T., O'CONNELL, D. N. Validation and cross-validation of a scale of self-reported personal experiences which predict hypnotizability. *Journal of Psychology,* 1962, **53,** 55–75.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Forms A and B.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale, Form C.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

(Received December 18, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 5, 390–394

EFFECT OF ENRICHED INPUT ON SCHIZOPHRENIC ABSTRACTION

ROY M. HAMLIN, H. CARL HAYWOOD,[1] AND ANGELA T. FOLSOM

Veterans Administration Hospital, Danville, Illinois, and University of Illinois

Closed ward, open-ward, and former schizophrenic patients were compared with nonschizophrenic patients. Ss gave abstract interpretations for single proverbs and for sets of 3 proverbs, the proverb sets providing an enriched input condition. Schizophrenics with mild and medium degrees of psychosis showed improvement in abstract responses with enriched input. Nonschizophrenic and severely schizophrenic Ss showed no improvement. When input was adequate, deficit in abstract responses disappeared entirely for the mildly schizophrenic Ss. Severely schizophrenic Ss showed deficit on all tests.

Although many studies show that schizophrenic patients perform at a significantly lower level than do nonschizophrenic subjects on abstracting tasks, some authors (Cameron, 1939; Hunt & Cofer, 1944; Lothrop, 1961) suggest that it is not abstracting ability as such that is lost. The apparent loss in abstract behaviors may be accounted for by a deficit in the communication process, either in information input or in the verbal expression of abstract concepts.

In a previous study, Blaufarb (1962) enriched input. He presented sets of proverbs to his subjects, each set containing three different (concrete) proverbs, but all three calling for a single abstract interpretation. He compared performance on these proverb sets with performance on single proverbs. Since the proverbs in a set were each (concretely) unique, the procedure involved something more than mere stimulus repetition. The task message was enriched in meaning or presented redundantly.

Blaufarb's study indicates that stimulus enrichment improves schizophrenic abstract responses but has *no effect* with normal subjects. This result suggests that schizophrenics do have an input deficit. The schizophrenic subjects, however, tended to score lower on both abstract tasks, on single proverbs, and (very slightly) on proverb sets. That is, the schizophrenics did show an input deficit, but perhaps also a primary loss in abstracting ability. Blaufarb justifiably emphasizes the input deficit and concludes that "chronic schizophrenics apparently have not lost the capacity for verbal abstraction, but may possibly fail to deal effectively in some ways with incoming stimuli [p. 475]." This conclusion would be strengthened by demonstrating that input enrichment could remove the abstracting deficit entirely.

The present study extends Blaufarb's results over four degrees of pathology—three schizophrenic groups, and one nonschizophrenic group. While the strategy of showing a difference between schizophrenic patients and nonschizophrenic controls has enjoyed a great deal of popularity, and has, in fact, yielded interesting information on the nature of schizophrenia, recent studies (Hamlin & Jones, 1963; Hamlin & Nemo, 1962; Haywood, Goldman, & Beck, 1961; Haywood & Moelis, 1963) have demonstrated the efficacy of sampling at several points along the dimension of severity of psychopathology. When patients are compared with other patients as well as with normal subjects, it is possible to draw inferences concerning the *process* of mental disorder as it exists in varying degrees of symptom manifestation and as it changes from acute onset through the course of treatment to recovery or to continued treatment.

In the present study, it was anticipated that Blaufarb's input enrichment would have different effects at the four levels of pathology considered. Specifically, only patients with a medium degree of schizophrenic pathology would benefit from enriched input. Former patients, representing a mild degree of pathology, would resemble Blaufarb's normal controls, as in a previous study of vocabulary

---

[1] Now at George Peabody College, Nashville, Tennessee.

(Hamlin & Jones, 1963). These schizophrenics in remission would not gain from single proverb scores to proverb set scores, as they would be subjects who had never lost, or who had recovered, input efficiency. Severely schizophrenic patients would not gain for the opposite reason: input disruption is too severe and is associated with generalized cognitive disorganization. The mild enrichment of the multiple-proverb procedure would be insufficient to overcome the deficit.

By including several degrees of psychotic disorganization, it was hoped that a mild level of pathology could be found where schizophrenics show no abstracting deficit but continue to show input deficit. Such a result (with obvious reservations) would support the possibility that input deficit is *primary*, and that abstracting deficit is a *secondary* consequence of the more fundamental input deficit.

METHOD

All subjects were male veterans, currently or formerly hospitalized in the Veterans Administration Hospital, Danville, Illinois. Four groups of 14 subjects each, equated for age and education, represented: (*a*) closed ward schizophrenics, (*b*) open-ward schizophrenics, (*c*) former schizophrenic patients in remission, and (*d*) nonpsychotic patients. The several studies mentioned previously, employing levels of pathology, indicate that these samples represent distinct steps in a continuum of schizophrenic pathology, and three of the four subject sources show highly reliable differences on Lorr's Psychotic Reaction Profile (unpublished research, Haywood, 1962). Since Blaufarb used nonpatient employees for his nonschizophrenic group, it was decided that a different and varied nonschizophrenic group should be used in the present study. The 14 subjects in the present nonschizophrenic group consisted of 6 neurotics, 5 personality disorders, and 3 medical patients without neuropsychiatric diagnosis. In both Blaufarb's and the present experiment, nonpsychotics showed very similar performances, so that the results reported cannot be attributed to hospitalization or to psychiatric disorders other than psychosis. Schizophrenic pathology would seem to be the relevant group-difference variable.

Subjects were given the oral vocabulary from the Wechsler Adult Intelligence Scale; 17 single proverbs suitable for abstract interpretation; and 17 sets of three proverbs each, all three proverbs in a set being suitable for a single abstract interpretation (cf. Blaufarb, 1962, for more detail). In each group, half of the subjects were given the 17 single proverbs first, and half were given the proverb sets first. The two proverb conditions were given approximately 5 days apart. Standard scoring was used for the vocabulary test. The proverbs responses were scored according to the system developed by Kaufman (1960), and described by Blaufarb (1962), who reports scorer reliability of .89. Scoring was done without knowledge of experimental condition or pathology level. The maximum possible score on any one set of 17 proverbs is 102, well above the highest score obtained by any subject.

TABLE 1

MEAN AGE, SCHOOL GRADE, AND VOCABULARY FOR THE FOUR GROUPS

Group	CA	School grade	Vocabulary[a]
I Closed ward	39.1	9.4	25.3
II Open ward	38.6	10.3	41.9
(Blaufarb: patients)			(42.3)
III Former patients	40.4	10.6	45.4
IV Nonschizophrenic	41.0	10.3	44.9
(Blaufarb: controls)			(41.3)

[a] Analysis of variance for vocabulary: $F = 9.98$, $p < .001$.

RESULTS

Table 1 shows the mean school grade completed, mean chronological age, and mean vocabulary score for each of the four subject groups. Differences in education and age are not significant. As anticipated, vocabulary scores increase significantly ($p < .001$) as psychopathology becomes less severe across the groups. The schizophrenics in remission show no deficit in vocabulary, as compared with the nonschizophrenic subjects. These results simply confirm the previous research of Hamlin and Jones (1963). Mean vocabulary scores for Blaufarb's two groups are included in Table 1. The slightly lower mean vocabulary for his nonschizophrenic group reflects his selection of subjects: he *matched*

TABLE 2

MEAN PROVERB SCORES FOR FOUR GROUPS UNDER TWO CONDITIONS OF PRESENTATION

Condition	Group I	Group II	Group III	Group IV
Single proverbs				
M	35.2	45.8	57.6	63.9
SD	15.35	9.99	13.57	11.80
Proverb sets				
M	34.9	52.4	62.7	61.9
SD	16.86	7.24	11.50	9.96
t	−0.18	3.43*	3.22*	−0.70

* $p < .01$, 2-tailed test.

TABLE 3

ANALYSIS OF VARIANCE OF PROVERB SCORES

Source	*df*	*MS*	*F*	*p*
Levels of pathology A	3	4,490.643	15.024	<.001
Orders of presentation B	1	6.036		
A × B	3	127.821		
Error (a)	48	298.895		
Proverb conditions C	1	155.571	6.186	<.05
A × C	3	120.643	4.797	<.01
B × C	1	60.036	2.387	<.20
A × B × C	3	19.821		
Error (b)	48	25.145		
Total	111			

Groups compared	I–II	I–III	I–IV	II–III	II–IV	III–IV
Single proverbs, *t*	2.17*	4.09**	5.55**	2.62*	4.38**	1.31
Proverb sets, *t*	3.56**	5.08**	5.14**	2.84**	2.89**	−0.20

* $p < .05$, two-tailed.
** $p < .01$, two-tailed.

his schizophrenic and control subjects on vocabulary.

The intercorrelations among the three measures used in the present study are: vocabulary-single proverbs, .66; vocabulary-proverb sets, .74; and single proverbs-proverb sets, .90. The high correlation of .90 between single proverbs and sets suggests the adequate reliability of the proverbs measures. This correlation may be high in part because of the wide range of pathology included in the present study.

Table 2 gives the mean scores for the four groups on single proverbs and on proverb sets. These scores show a consistent tendency to run slightly higher than Blaufarb's proverb scores. This constant error has no appreciable effect on the results reported, and is a result presumably of the scoring difficulty noted by Blaufarb: "minor additional scoring criteria had to be devised [p. 473]."

As can be seen from Table 2, the two less psychotic schizophrenic groups show significant gain from single proverb scores to proverb set scores. The most pathological schizophrenic group, and the nonschizophrenic group, show no tendency to such a gain. The performance of the schizophrenics in remission is of special interest. These subjects show significant gain, interpreted as evidence of input deficit. When this input deficit is corrected by the enriched stimulus procedure, there is no evidence whatsoever of abstracting deficit. The mean score on proverb sets is a fraction of a point higher than for the nonschizophrenic group. Table 2 shows also, however, that the schizophrenics in remission are not significantly different from the nonschizophrenics on either proverbs test; nor do they show deficit on vocabulary (Table 1). The two more severely schizophrenic groups show significant deficit on both proverbs tests. The enriched input procedure fails to change the general pattern of relative deficit.

Table 3 presents the basic analysis of variance (trend analysis) for the data, paral-

TABLE 4

MEAN AGE, SCHOOL GRADE, VOCABULARY, AND PROVERBS SCORES FOR 24 SUBJECTS EQUATED FOR VOCABULARY

Group	CA	School grade	Vocabulary	Single proverbs	Proverb sets	Proverbs gain
I Closed ward	40.8	9.5	36.3	45.7	47.7	2.0
II Open ward	36.5	9.7	36.0	40.3	48.5	8.2
III Former patients	41.2	10.7	35.5	50.3	60.3	10.0
IV Nonschizophrenic	41.5	10.2	36.5	60.7	58.8	−1.8

leling Blaufarb's analysis. The significant Groups × Conditions interaction indicates that the two less schizophrenic groups showed greater improvement than the other two groups in their abstracting performance when presented with the proverb sets.

The focal result indicated by Table 3 cannot be attributed to initial differences in vocabulary scores. In Blaufarb's study, schizophrenics and controls were equated for vocabulary. In planning the present study, matching vocabulary across a range of pathology was not considered feasible. As it turned out, overlap in vocabulary is sufficient to permit an analysis with vocabulary equated for each pathology level. From the 56 subjects in the study, it was possible to select 24, with each cell of the design proportionally represented, and with vocabulary comparable for every level. No basis for selection other than vocabulary was involved. Table 4 shows the close match with vocabulary, and gives mean chronological age, education, and gain scores from single proverbs to proverb sets.

Table 5 presents the analysis of variance for these 24 subjects. As expected, the highly significant difference for total proverbs scores ($p < .001$) largely disappears ($p < .20$). That is, vocabulary and proverbs scores are correlated, as already reported. The Groups × Conditions interaction, however, is not affected ($p < .01$), and indicates that the medium and mild schizophrenic groups showed greater improvement when presented with proverb sets. Pathology level, and not vocabulary, would seem to be the relevant variable.[2]

TABLE 5

ANALYSIS OF VARIANCE OF PROVERB SCORES FOR 24 SUBJECTS EQUATED FOR VOCABULARY

Source	*df*	*MS*	*F*	*p*
Levels of pathology A	3	625.139	1.91	.20
Orders of presentation B	1	114.083		
A × B	3	180.583		
Error (a)	16	326.917		
Proverb conditions C	1	252.083	15.05	.01
A × C	3	90.028	5.38	.01
B × C	1	30.083	1.80	.20
A × B × C	3	28.583	1.71	
Error (b)	16	16.750		
Total	47			

In fact, the data suggest that the medium and mild schizophrenic subjects outgain the nonschizophrenics on proverb sets at every vocabulary level, and at every level of single proverbs score. (The closed ward schizophrenics cannot be compared because of their generally low test performance.) Table 6 indicates a *relative* gain of seven or eight points at each vocabulary level, and small *relative* gains at each single proverbs level.

TABLE 6

MEAN PROVERBS SCORES FOR 28 SCHIZOPHRENIC AND 14 NONSCHIZOPHRENIC SUBJECTS, BY VOCABULARY AND BY SINGLE PROVERBS LEVELS

Level	Schizophrenics				Nonschizophrenics				Relative gain
	N	Single proverbs	Proverb sets	Gain	*N*	Single proverbs	Proverb sets	Gain	
Vocabulary									
20–39	10	45.4	50.2	4.8	4	58.5	56.0	−2.5	7.3
40–49	9	49.6	57.9	8.3	5	61.2	61.6	.4	7.9
Above 49	9	60.9	65.4	4.5	5	71.0	67.0	−4.0	8.5
Single proverbs									
30–39	8	35.6	45.9	10.3	0				
40–49	2	44.5	52.5	8.0	2	43.0	48.5	5.5	2.5
50–59	11	54.3	59.5	5.2	3	56.7	59.3	2.6	2.6
60–69	4	62.0	64.8	2.8	2	62.0	60.0	−2.0	4.8
Above 69	3	76.3	75.3	−1.0	7	73.6	67.4	−6.2	5.2

---

[2] An analysis of gain scores, using all 56 subjects, and with covariance adjustment for differences in vocabulary confirms the effect of enriched input ($p < .01$). The authors wish to express appreciation to John C. Stumpf and Jack M. Becktel (undated letter), of the Veterans Administration Midwest Research Support Center, who further confirm the results in a more sophisticated analysis of the data, showing that the assumption for covariance is met, and that a covariance analysis is not necessary.

DISCUSSION

The experiment confirms Blaufarb's results, and in general, the predictions made for the several degrees of pathology. Under enriched stimulus conditions, the subjects with medium and mild degrees of schizophrenic pathology show significant improvement in verbal abstracting performance. Nonschizophrenic subjects show no improvement. The least schizophrenic subjects give just as good abstract responses as nonschizophrenics to the multiple proverbs. As predicted, the severe closed ward schizophrenics show impairment on all tasks, and are unable to benefit from the enriched stimulus procedure.

As Blaufarb argues, the multiple proverbs procedure does not provide abstract responses for the subject, nor make the abstracting task easier. If either of these possibilities were the case, the nonschizophrenic subjects would gain. What the enriched stimulus procedure does seem to do is overcome an input deficit. A concrete interpretation is superficially appropriate for a single proverb, but three different concrete proverbs make almost any concrete interpretation logically inadequate. With this restriction on response choice, the schizophrenic shows increased evidence that he has retained abstracting capacity. The enrichment of input may consist largely in forcing upon the subject the subtle intent of the instructions: "give me the meaning or idea behind each proverb."

The results of the study are not inconsistent with the possibility that input deficit is more fundamentally characteristic of schizophrenia than is loss of abstracting ability. Impaired input may constitute a primary deficit, which then leads to abstracting deficit. Blaufarb makes clear his definition of abstracting capacity: under usual task conditions, the schizophrenic "may be unable to relate the task stimuli to the task instructions and/or be unable to relate the task stimuli to the abstract solution which is apparently available to him [p. 474]." If abstract solutions are still available, abstracting capacity is "intact." This availability of abstract solutions is not adequately demonstrated either for the open ward schizophrenics or, especially, for the closed ward patients. These groups show significant deficit on both proverbs procedures. Yet it could be argued that input deficit accounts for *apparent* abstracting deficit in mildly schizophrenic subjects; and then leads to actual abstracting deficit, and to generalized intellectual disorganization, in more severe degrees of pathology.

At best, the studies only suggest such a rationale. The enriched input procedure fails to change materially the general pattern of deficit across the levels of pathology considered. On the other hand, the multiple proverbs are a rather mild approach to improving input; and the results are encouragingly consistent across a range of pathology and competence levels. The studies should encourage special research attention to the role and nature of input deficit in schizophrenia.

REFERENCES

BLAUFARB, H. A demonstration of verbal abstracting ability in chronic schizophrenics under enriched stimulus and instructional conditions. *Journal of Consulting Psychology*, 1962, **26**, 471–475.

CAMERON, N. Deterioration and regression in schizophrenic thinking. *Journal of Abnormal and Social Psychology*, 1939, **34**, 265–270.

HAMLIN, R. M., & JONES, R. E. Vocabulary deficit in improved and unimproved schizophrenic subjects. *Journal of Nervous and Mental Disease*, 1963, **136**, 360–364.

HAMLIN, R. M., & NEMO, R. S. Self-actualization in choice scores of improved schizophrenics. *Journal of Clinical Psychology*, 1962, **18**, 51–54.

HAYWOOD, H. C., GOLDMAN, JACQUELINE R., & BECK, SALLY B. Length of hospitalization and Rorschach form responses of hospitalized schizophrenics. *Perceptual & Motor Skills*, 1961, **13**, 269.

HAYWOOD, H. C., & MOELIS, I. Effect of symptom change on intellectual function in schizophrenia. *Journal of Abnormal and Social Psychology*, 1963, **67**, 76–78.

HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 971–1032.

KAUFMAN, LEE N. The development of a proverb scale for the measurement of thinking pathology and a further investigation of the process-reactive dimension. Unpublished master's thesis, University of Illinois, 1960.

LOTHROP, W. W. A critical review of research on the conceptual thinking of schizophrenics. *Journal of Nervous and Mental Disease*, 1961, **132**, 118–126.

(Received December 22, 1964)

Volume 70 December 1965 Number 6

Journal of Abnormal Psychology

Editor: Howard F. Hunt, *New York State Psychiatric Institute and Columbia University*

Associate Editor: William N. Thetford, *College of Physicians and Surgeons, Columbia University*

Consulting Editors

Theodore X. Barber
Medfield Research Foundation

Harold Basowitz
New York University

Eugene I. Burdock
New York State Psychiatric Institute and Columbia University

John M. Butler
University of Chicago

Arthur C. Carr
New York State Psychiatric Institute and Columbia University

Loren J. Chapman
Southern Illinois University

John H. Cumming
New York State Department of Mental Hygiene

W. Grant Dahlstrom
University of North Carolina

Jerome D. Frank
Johns Hopkins University

Sol L. Garfield
Columbia University

Ann Magaret Garner
Lincoln, Nebraska

Israel Goldiamond
Institute for Behavioral Research

Harrison G. Gough
University of California, Berkeley

Starke R. Hathaway
University of Minnesota

Ralph Heine
University of Chicago

Ernest R. Hilgard
Stanford University

Lawrence C. Kolb
New York State Psychiatric Institute and Columbia University

Sheldon J. Korchin
University of California

William G. Lawlor
Fordham University

Ardie Lubin
U. S. Naval Hospital, San Diego

Salvatore Maddi
University of Chicago

David Rosenthal
National Institute of Mental Health

Alan K. Rosenwald
Illinois State Psychiatric Institute

Saul Rosenzweig
Washington University, St. Louis

Jacques Rutschmann
New York State Psychiatric Institute

Seymour Sarason
Yale University

Joseph Schachter
New York State Psychiatric Institute and Columbia University

Roy Schafer
Yale University

John Shlien
University of Chicago

Morris I. Stein
New York University

Charles Strother
University of Washington

Harold L. Williams
University of Oklahoma School of Medicine

C. L. Winder
Michigan State University

Joseph Zubin
New York State Psychiatric Institute and Columbia University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Abnormal Psychology* is devoted to basic research and theory in the broad field of abnormal behavior, its determinants, and its correlates. The following general topics fall within its area of major focus: (*a*) psychopathology—its development, treatment or remission, and its symptomatology, and course; (*b*) normal processes in abnormal individuals; (*c*) pathological or atypical features of the behavior of normal persons; (*d*) experimental studies, with human or animal subjects, relating to emotional behavior or pathology; (*e*) social or group effects on adjustment and pathological processes; (*f*) tests of hypotheses from psychoanalytic or other psychological theory. Thus, case histories, experiments on hypnosis, theoretical papers of scholarly substance on personality or emotional abnormality, studies of patient populations, analyses of abnormal behavior and motivation in terms of modern behavior theories, studies of therapy and behavior change, to cite a few illustrative items, all would fall within the boundaries of the Journal's interests.

The JOURNAL is issued bimonthly, one volume per year. The subscription price is $10.00 annually (foreign, $10.50). Single issues are $2.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor:

Howard F. Hunt
Journal of Abnormal Psychology
Department of Psychology
N. Y. State Psychiatric Institute & Columbia University
722 West 168th St.
New York, N. Y. 10032

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or may be obtained from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

ELIZABETH S. REED
Advertising Manager

FRANCES L. BREWER
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Entry as second-class matter at Lancaster, Pa. pending.
Copyright © 1965 by the American Psychological Association, Inc.

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 395–402

DESENSITIZATION, SUGGESTIBILITY, AND PSEUDOTHERAPY[1]

PETER J. LANG
University of Wisconsin

A. DAVID LAZOVIK
University of Pittsburgh

DAVID J. REYNOLDS
Veterans Administration Hospital, Leech Farm Road, Pittsburgh

44 snake phobic Ss participated in laboratory experiments assessing the degree of fear change associated with systematic desensitization, no treatment, placebo treatment, and the trait of suggestibility. Desensitization Ss showed significantly greater fear reduction than controls, while placebo Ss changed no more than did untreated Ss. Successful desensitization was relatively independent of suggestibility. Desensitization of specific fears generalized positively to other fears, and among desensitization Ss, degree of fear change could be predicted from measurable aspects of therapy process.

Positive change in psychotherapy has often been attributed to the relationship established between therapist and client, in the context of treatment. This phenomenon has been variously described as transference, suggestibility, the "hello-goodbye," or placebo effect. Lang and Lazovik (1963) demonstrated that phobic subjects who were briefly exposed to desensitization psychotherapy showed a significantly greater reduction in fear behavior than did untreated controls. The present experiment is an extension of this work, specifically to determine if the obtained change can be assigned to placebo effects.

A psychotherapeutic placebo is not as readily developed as the control medication of drug research. Psychological treatment is usually more prolonged; it involves a more complex interpersonal relationship, and the distinction between a placebo and the clinically effective agent can be less clearly delineated. Furthermore, the institution of a long pseudotreatment is generally impractical in the clinical situation. Not only is it wasteful of the patient's and therapist's time, but the necessary deception and delay of legitimate treatment can seldom be justified, and may be detrimental to the patient's future chances for cure.

These objections had considerably less validity in the experimental context considered here, which offered an opportunity to employ a stringent placebo control group. The subjects used in this research displayed phobic reactions to nonpoisonous snakes of considerable intensity. They systematically avoided snakes or places such as zoos or camping trips where they might be found. These individuals behaved so "unrealistically" as to react with anxiety not only to snakes themselves, but to pictures of snakes, artifacts (a snakeskin belt), similar shapes, or the mere mention of the word. Nightmares concerning snakes were not uncommon in this group. Such statements, descriptive of their fear, as the following have been recorded: "There's nothing in the world that I'm more afraid of than snakes." "I'm just weak when I see one." "I don't even want to look at them in books." In nearly all respects, except the degree to which the phobia is an omnipresent source of concern, this fear resembled those that patients themselves bring to the attention of psychiatrists and psychologists.

However, these subjects were not drawn

---

[1] This research was supported by Public Health Service Grant M-3880, from the National Institute of Mental Health. The authors would like to thank J. Geer, R. Miller, R. Romano, and Jean Wilkinson, who participated as therapists in this project. Appreciation is also expressed to Lynne Norris and R. Wiater, who assisted in organizing the data collection and analysing the results.

from a clinical population, but were selected from among normal college students by a classroom questionnaire and subsequent interview (their frequency is approximately one or two per hundred students surveyed). They volunteered to participate in an experiment on psychotherapy. They were instructed that the procedures used might reduce or eliminate their fear, but that the main purpose of the project was a scientific evaluation of different therapy methods.

The advantages of working with this population are obvious. Treatment can be specific. The number of sessions may be arbitrarily controlled. The moral responsibility to choose the best treatment is not involved. As this phobia is not a central life problem and these subjects were generally more stable than patients, extratherapeutic incidents less frequently interfere with the therapy process. Most important, the necessary rigor of experimental procedure can be closely respected.

As in all translations of natural events into a laboratory context, something is lost in vivacity. In this case, some of the subject characteristics mentioned above, the rigidity of procedure, are obviously different from what is characteristically found in the consulting room. However, much can be gained in the exactness and clarity with which natural events are elucidated. Thus, the present study does not constitute a clinical test of desensitization therapy (clinical statistics have already been reported by Wolpe, 1958), but it is rather an attempt to illuminate through experiment the mechanism by which fear reduction is achieved in this method. Specifically, this research is designed to determine if placebo effects account for the positive results achieved by systematic desensitization therapy, and also to evaluate the overall contribution of suggestibility, as a personality trait, to progress in desensitization and posttherapy fear reduction.

METHOD

Desensitization Therapy

The procedure has been described in detail elsewhere (Lang & Lazovik, 1963; Lazovik & Lang, 1960). Each subject first experienced 5 training sessions. At this time an *anxiety hierarchy* was constructed—a series of 20 phobic situations graded from least to most frightening. The subject was also trained in hypnosis and deep-muscle relaxation (Jacobson, 1938). Subsequently he participated in 11 desensitization sessions. During these meetings he was hypnotized and instructed to vividly imagine the scenes described in the hierarchy.[2] The scenes were presented one at a time, starting with the least frightening, and repeated until the subject no longer reported coincident anxiety. This desensitization occurred in the context of deep-muscle relaxation, which is held to counter condition or "reciprocally inhibit" (Wolpe, 1958) the fear response. Sessions lasted approximately 45 minutes at the rate of 1 or 2 per week.

Pseudotherapy

An effort was made to involve the subject in a treatment procedure which was therapeutically neutral except for the therapist-client relationship. Because desensitization was to be evaluated, all procedures employed in that method were included in pseudotherapy. The subjects first experienced the same 5 training sessions as in desensitization, followed by 11 pseudotherapy sessions. The procedure for each of these latter sessions was the same. The subject was first hypnotized and then told to relax deeply. During the first 15 minutes of the therapy hour the subject was asked to imagine a series of scenes, which he had previously described as pleasant and relaxing. The last half hour was keyed to the hierarchy items. The items were taken in order during the course of therapy, and provided starting points for a discussion of nonanxiety evoking aspects of the subject's life. The therapist generally behaved in a nondirective manner. However, he did attempt to prevent phobic responses from being made in the context of the therapy hour. It was held that the occurrence of phobic verbal behavior could lead to positive change, for the same reason that theory predicts it in desensitization, that is, inhibition of anxiety by relaxation or comfort responses instigated by the therapist. Thus, the pseudotherapist gently steered the conversation away from phobic or other sensitive material—in the main, by reinforcing nonsensitive topics with nods and verbal signs of his attention, and failing to similarly reinforce comments directly pertinent to the subject's fear. For example, if the hierarchy item concerned a black snake at the farm of an uncle, the developing conversation might concern farms or farm animals, or experiences with the uncle, whatever the subject spontaneously brought up that did not directly refer to his anxieties.

The theoretical orientation given these subjects facilitated the pseudotherapist's task. In the first session the following explanation of procedure was offered:

> Perhaps if I impart a little bit of psychological theory to you, it will help you understand our

---

[2] Six recently treated subjects experienced desensitization without hypnosis.

> next procedure. You have learned how to relax your muscles and this is going to be important in dealing with your fear. Previous research suggests that people with fears like yours have higher levels of autonomic tonus than others. What this means is that all the vegetative systems, digestive, circulatory, as well as the neural and muscular systems tend to overreact. They overreact mainly because they actually start from a higher level of tonus. Some psychologists hold that if this generally high tonus level could be reduced, the way would be paved for a general reduction in specific fears.
>
> What we hope to do in the rest of our sessions together is help you achieve a general lowering of the tonus levels. We will do this by first having you relax as you have learned to do. Then you will be hypnotized and even deeper relaxation will be achieved. To facilitate this process we will suggest pleasant scenes, from time to time, for you to concentrate on.
>
> As you probably know, fears such as yours are often related to situations which seem unimportant. After you are comfortable, that is deeply relaxed, we will want to begin a discussion with some of the items you used in the fear hierarchy. We will use these as starting points in our talks but plan to deviate from them. Our goal is to explore a number of areas of living, and produce a lowering of your overall tension levels. The theory holds that a person may overcome his fear if he obtains a better understanding of himself and learns to deeply relax.

The subjects participated eagerly. They came regularly and none discontinued treatment before the end of the experiment. Furthermore, no subject reported that he suspected he was involved in a nonviable treatment. The therapists reported consistently close, empathic relationships with these subjects, comparable or superior to those achieved during desensitization.[3]

Subjects

This research has been underway continuously since September 1960, and the present report describes the total sample of 44 subjects. Included are 23 desensitization subjects, 11 untreated controls, and 10 subjects who participated in pseudotherapy.[4]

All subjects were introductory psychology students at the University of Pittsburgh. They rated their fear of nonpoisonous snakes as "intense," on a fear questionnaire, and were included in this research only if a psychological interview corroborated this statement. The subjects who appeared to have impairing physical disabilities or latent psychosis (based on the psychotic scales of the MMPI or the clinical judgment of the interviewer) were excluded. None of the subjects in this study were being seen elsewhere because of psychological problems.

Assignment to groups was essentially random, although some pretreatment effort to balance control variables was made. A more elaborate description of the selection battery has already been reported (Lang & Lazovik, 1963).

Procedure

After the selection tests and interviews, all subjects were administered the Stanford Hypnotic Susceptibility Scale Form A (SHSS; Weitzenhoffer & Hilgard, 1959). Subsequently all subjects except the untreated controls participated in the five therapy training sessions, after which SHSS Form B was administered. Fear intensity was measured both before and after the subsequent 11 desensitization or pseudotherapy trials. Untreated subjects were not seen, except for evaluation sessions.

The measures employed (Lang & Lazovik, 1963) were: The Fear Survey Schedule (FSS), FSS Number 38, an avoidance test, and the "fear thermometer." The FSS is a list of 50 phobias that subjects rate on a 7-point scale. Item Number 38 is the snake item. In the avoidance test the subject was confronted with an alive, tame blacksnake. He was invited to approach the animal, in a controlled setting. The closest point of approach to the animal provided the basis for his test score. If the subject held the animal, he achieved a score of 1, refusal to go to the test room and observe the snake yielded a score of 19. Immediately after this experience the subject rated his situational anxiety during the avoidance test on a 10-point "fear thermometer" (Walk, 1956). An open-ended fear interview provided additional qualitative information.

The experimenters in this research were all experienced psychotherapists. Most of the subjects were seen by two psychologists who have a full time psychotherapy practice. The same experimenters who saw the pseudotherapy subjects also treated desensitization subjects. Fear evaluation was conducted by an experimenter who participated in no other aspect of the procedure.

---

[3] Two factors probably helped considerably to insure that the pseudotherapist was not unmasked: (*a*) treatment was relatively brief, (*b*) metropolitan residents were being treated for a snake phobia. Therefore, subjects had little or no contact with the phobic object outside of the experimental situation. If subjects were treated for a more omnipresent fear over a longer time period, unfavorable feedback from the life situation might make the deception much more difficult to maintain.

[4] The untreated subjects and 13 desensitization subjects were described in a previous report (Lang & Lazovik, 1963). All desensitization subjects were combined as there was no significant difference between this 1963 group and subjects who have since participated, either in selection method or results on the major assessment variables. Change scores for the 1963 sample and all other subsequent desensitization subjects were, respectively: avoidance test, .23 and .32; fear thermometer, 2.47 and 2.38; FSS Number 38, 1.38 and 1.44.

TABLE 1

MEAN PRE- TO POSTTREATMENT CHANGE SCORES FOR ALL FEAR MEASURES

Group	Avoidance test	Fear thermometer	FSS Number 38	Fear survey
Combined control ($N = 21$)	−.03	1.14	.48	12.14
Pseudotherapy ($N = 10$)	.14	1.30	.40	12.50
No treatment ($N = 11$)	−.19	1.00	.54	11.82
Desensitization ($N = 23$)[a]	.27	2.43	1.41	18.64
(hierarchy items completed)				
15 or more ($N = 13$)	.47	3.58	2.31	23.77
Less than 15 ($N = 10$)	.01	.89	.11	7.33

Note.—The avoidance test score is the percentage change statistic previously described (Lang & Lazovik, 1963). All other change scores were simply the difference between pre- and posttests. The correlations between initial performance and fear change for all measures were insignificant and inconsistent in direction.

[a] Data were incomplete for three *S*s. The Fear Thermometer and FSS *N*s are 22 and 21, respectively.

RESULTS

Mean fear behavior change scores from pretest to posttest are presented in Table 1. The similarity of the pseudotherapy and no-treatment means is readily apparent. The *t* tests yielded no significant differences between these groups for any measure of fear change (see Table 2). Furthermore, the change score frequency distributions for the two groups were normal and overlapping. Pseudotherapy and no-treatment subjects were therefore combined into a single control group for subsequent analyses.

Mean change scores for the combined control group and the desensitization group are presented in Table 1. Desensitization subjects showed significantly greater fear reduction than controls as measured by all three indices of snake phobic behavior. The *t*'s are presented in Table 2 and desensitization and control frequency distributions are illustrated in Figure 1. Except for the avoidance test the control sample yielded essentially normal distributions of fear change. The desensitization subjects have a primary modal score in the same interval as the controls, but the distributions are skewed positively or are frankly bimodal.

Progress in systematic desensitization is directly measured by the portion of the 20-item anxiety hierarchy completed in the 11 therapy sessions. Table 3 reveals that all measures of fear reduction are positively related to the number of hierarchy items successfully completed by each subject. No pretest measure of snake phobic behavior correlated significantly with the number of items subsequently completed, although the first presentation of the FSS and items completed yielded an r of $-.435$, $(p < .05)$. Number of items completed was wholly unrelated to any of the SHSS measures (see Table 3).

A previous review of part of these data (Lang & Lazovik, 1963) suggested that while subjects who completed 15 items or more showed considerable positive change, subjects completing less than 15 showed no more change than untreated subjects. The current

TABLE 2

The *t* TESTS OF MEAN FEAR CHANGE SCORES FROM PRE- TO POSTTREATMENT

Group	Avoidance test	Fear thermometer	FSS Number 38	Fear survey
Combined control versus desensitization	2.57*	2.12*	2.19*	1.25
Combined control versus 15 or more	3.26**	3.44**	3.99***	2.52*
Combined control versus less than 15	.14	.41	1.85	.41
Less than 15 versus 15 or more	2.33*	3.28**	5.00***	2.26*
Pseudotherapy versus no treatment	1.67	.48	.58	.12

Note.—The desensitization group was subdivided into two groups on the basis of performance in therapy. "15 or more" and "less than 15" refers to the number of anxiety hierarchy items successfully completed by the group during desensitization.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

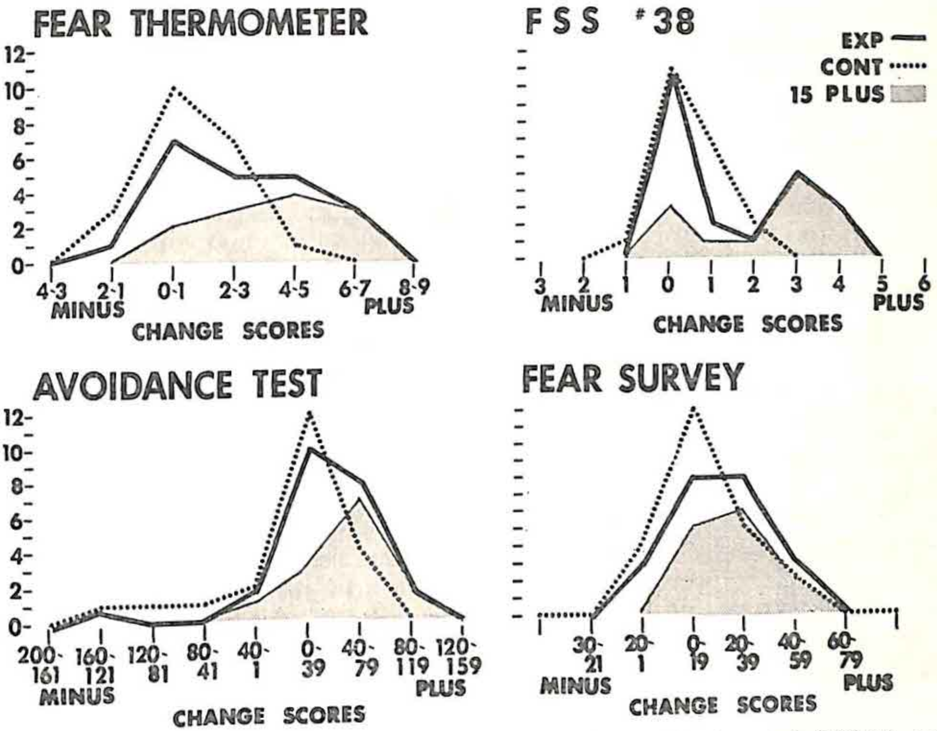


FIG. 1. Fear change score frequency distributions for the combined control (CONT) and desensitization (EXP) samples. The shaded areas of the desensitization groups' distributions define the subjects who successfully completed 15 or more anxiety hierarchy items.

enlarged sample confirmed this trend. It may be noted in Table 1, that subjects completing more than 15 items have the highest average change scores for all fear measures. Furthermore, Figure 1 shows that the 15 plus subjects clearly account for the positive skew of the distributions of experimental subjects. If the 15 plus group is eliminated, the remaining subjects are distributed normally around the control mode. These impressions were assessed by the *t* test, and the results are reported in Table 2: 15 plus subjects show significantly more fear reduction than either controls or less than 15 subjects. The latter group was not superior to the control group on any change measure.

The correlations between the pretest fear measures and SHSS Forms A and B, and the difference score, are at the top of Table 4. Inspection of these coefficients reveals that only the avoidance test is significantly related to suggestibility. Correlations between fear reduction and the SHSS measures are listed in the bottom half of Table 4. It may be noted that avoidance test change is unrelated to suggestibility for either the desensitization or control subject. For control subjects both the fear thermometer and FSS 38 change are positively related to measures of suggestibility, and the Fear Survey shows a tendency in this direction. There is no significant relationship between the desensitization group's SHSS scores and any fear change measure.

TABLE 3

PEARSON CORRELATIONS BETWEEN THE NUMBER OF ANXIETY HIERARCHY ITEMS SUCCESSFULLY COMPLETED DURING DESENSITIZATION AND MEASURES OF FEAR CHANGE AND THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE

Fear change	r	SHSS form	r
Avoidance test	.40*	A	−.06
Fear thermometer	.50*	B	−.03
FSS Number 38	.60**	Change	−.05
Fear survey	.50*		

Note.—SHSS change is the difference between Forms A and B.
* $p < .05$.
** $p < .01$.

TABLE 4

PEARSON CORRELATIONS BETWEEN THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE AND THE INITIAL MEASURES OF FEAR AND FEAR CHANGE SCORES

Fear measure	SHSS Form		
	A	B	Change
Total sample ($N = 44$)			
(Initial score)			
Avoidance test	−.34*	−.47**	−.06
Fear thermometer	−.14	−.22	.04
FSS Number 38	.10	.19	.11
Fear survey	.00	−.10	−.13
Control group ($N = 21$)			
(Change score)			
Avoidance test	.00	.12	.06
Fear thermometer	.33	.46*	.15
FSS Number 38	.47*	.48*	−.02
Fear survey	.29	.36	−.23
Desensitization group ($N = 23$)			
(Change score)			
Avoidance test	−.16	.11	.28
Fear thermometer	−.16	.18	.39
FSS Number 38	.07	.02	−.19
Fear survey	.08	.07	.02

Note.—All correlational statistics were computed on the IBM 7090 at the University of Pittsburgh Data Processing Center.

* $p < .05$.

** $p < .01$.

DISCUSSION

Desensitization subjects clearly showed a significantly greater decrease in phobic behavior than did controls; subjects whose therapeutic experience was restricted to a placebo relationship yielded no more positive change than did the untreated group.

These results imply that the reduction in fear following desensitization does not stem from a suggestion to change implicit in being "in therapy." Both placebo and desensitization subjects were asked to participate in a procedure which had fear reduction as a goal, and both groups spent equal time at this task. The findings also indicate that hypnosis, training in muscle relaxation, hierarchy building, and their continued use in a therapeutic context, do not in themselves produce change in fear behavior. Finally, the results suggest that the relationship which developed between the experimenter and subject, with possible transference effects, was not in itself the vehicle of change.

The above inferences clearly depend on the extent to which the therapists kept their communications about goals the same for all subjects, and how well they succeeded in maintaining interpersonal relationships of comparable warmth and intensity with their client-subjects under both desensitization and placebo conditions. A number of factors argue that a high degree of comparability was achieved. The formal, experimental setting in which this research was carried out certainly assisted in this task. Each step of the procedure was defined, and a considerable portion of the therapists' communications with the subject were read from a mimeographed program. Furthermore, it is unlikely that the therapists who saw most of the subjects in this study were personally committed to any outcome. They were paid for their participation, at the same rate per therapy hour regardless of the procedure administered, and they were involved in no other aspect of the experiment. It must also be mentioned that all subjects were seen by clinicians with considerable experience in interviewing and traditional psychotherapy, who would be well above average in their ability to meet the relationship requirements of the experiment. Finally, the posttherapy interviews with the subjects, and discussions with the therapists, did not suggest that placebo subjects were less closely involved, nor did they yield subjects who doubted the fact that a true therapeutic procedure was being undertaken.

On the other hand, the possibility of error in procedure can never be completely discounted, and is best detected by independent attempts at replication. Furthermore, despite the fact that many of the same procedures were used in both placebo therapy and desensitization, the way they were used undoubtedly influenced the character of the experimenter-subject relationship. Unlike the pseudotherapy group, desensitization subjects received regular feedback concerning progress through the hierarchy. Clearcut success experiences (as well as failures) were part of the therapeutic interaction. In this context, a good therapeutic relationship may be one in which the therapist has gained the properties of a reinforcer, and this capacity is in the service of a specific program of behavior change. Desensitization would then progress more rapidly if approval is given for the completion of items. The better the relationship, the more effective such reinforcement is likely to be. On the other hand, it should be pointed out that the completion of an item is in itself reinforcing and many subjects compliment themselves on their own progress. Thus, the quality of the relationship may be less relevant to a well-motivated subject's performance. These variables merit the continued attention of researchers.

While the desensitization subjects as a group showed considerable fear reduction, the change score distributions are skewed and some individual members changed no more than controls. The current experiment was limited to a brief 11 desensitization sessions, and it is, of course, possible that all subjects would have improved with a sufficient exposure. It is also possible that there are personality differences between those who profit and those who do not. As the sample size increases this contingency will be explored. In any event, suggestibility is clearly not one of the potential distinguishing traits. While a significant part of the control subject's fear change is attributable to the SHSS, no similar relationship between this scale and fear reduction was found for the experimental group. For desensitization subjects the positive effects of treatment were so overriding as to render undetectable any variance assignable to suggestibility.

All measures of fear change yielded high positive correlations with number of hierarchy items completed. Furthermore, subjects who completed at least 15 items showed significantly less fear than controls at the end of the experiment; subjects who failed to move this far along in the hierarchy were no different from pseudotherapy or untreated subjects. The correlations indicate that a general relationship exists between fear change and a measurable aspect of therapy process. It is unlikely that Item 15 represents a "critical point" in this process. However, given the measures used here, the data suggest that the therapeutic task must be well advanced before effects clearly greater than those achieved by control subjects are observed.

As would be predicted by an S-R theoretical model, fears other than the phobia treated showed less marked reduction. The desensitization group as a whole yielded significantly greater change than controls on FSS Number 38; an analogous result was not obtained for the entire FSS. Nevertheless, 15 plus subjects did show a significant reduction on the FSS when compared either to control subjects or to the less than 15 group. These findings suggest that the treatment was specific, as intended, and when it was successful that fear reduction generalized positively.

These findings are inconsistent with some predictions from psychoanalytic theory. Fear reduction occurred without exploration of the phobia's dynamic background. Posttreatment interviews revealed no examples of symptom substitution. Positive generalization of fear reduction was indicated, rather than an increase in other fears. Furthermore, a transference cure interpretation of the fear change is weakened by the failure to find marked placebo effects.

Although the relationship by itself was shown to be less effective therapy than client-centered theorists have sometimes argued, this experiment does not constitute a test of that treatment method. A quasi-non-directive technique was employed as part of the pseudotherapy procedure, but its object was the opposite of client-centered therapy, to avert affect-laden statements rather than to encourage them. As was already suggested, theory would predict positive change in true client-centered treatment for the same reasons as for desensitization. Advocates of the latter treatment hold, of course, that behavior therapy is more specific and systematic, and therefore faster and more thorough. The present experimental design offers an excellent setting for a test of this assumption. Now that the level of placebo changes is known the way is open for meaningful, comparative evaluation of psychotherapies.

Both the limitations and the advantages of the subject population used in this experiment were mentioned at the outset, and they must be continually held in mind. However, the course of laboratory desensitization very closely followed the treatment process reported by clinical workers. Considerable success was achieved in 11 desensitization sessions, which compares favorably with the 11.6 average sessions per hierarchy reported by Wolpe (1958) for neurotic patients. Furthermore, many of the phenomena observed by the authors in clinical cases, and reported by others, were found with these experimental subjects. The systematic weakening of anxiety with repeated presentation of a hierarchy item was of course the typical result. However, subjects ocassionally showed a perseveration of anxiety, and apparent summation with repetition, that presented all the difficulties that this situation creates when it occurs in a clinical case. Similarly, problems with visualization were frequently observed, as well as avoidance behavior when individual items were particularly upsetting. While these problems were undoubtedly fewer and less intense than would occur in clinical practice, they did not appear to differ in quality or kind.

Despite this apparent versimilitude, generalization to the clinic must be cautiously undertaken. Many issues are raised that need more intensive investigation. Nevertheless, the results obtained here encourage four important conclusions: (*a*) simply being in a therapeutic context and relating to a therapist (even when all the trappings of desensitization—hypnosis, hierarchy building, relaxation—are included) does not in itself effect the important changes in phobic behavior achieved by systematic desensitization; (*b*) successful desensitization is relatively independent of the subject's suggestibility; (*c*) change in verbal and motor indices of fear behavior may be directly predicted from measurable events that occur during desensitization therapy; (*d*) the systematic desensitization of a specific fear generalizes positively to other fears. These findings are further presumptive evidence that desensitization therapy process conforms to its theoretical model, that it is an adapting out or systematic counterconditioning of fear responses to specific stimuli.

REFERENCES

JACOBSON, E. *Progressive relaxation*. Chicago: Univer. Chicago Press, 1938.

LANG, P. J., & LAZOVIK, A. D. Experimental desensitization of a phobia. *Journal of Abnormal and Social Psychology*, 1963, **66**, 519–525.

LAZOVIK, A. D., & LANG, P. J. A laboratory demonstration of systematic desensitization psychotherapy. *Journal of Psychological Studies*, 1960, **11**, 238–247.

WALK, R. D. Self-ratings of fear in a fear-invoking situation. *Journal of Abnormal and Social Psychology*, 1956, **52**, 171–178.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale*. Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WOLPE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford Univer. Press, 1958.

(Received August 17, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 403–404

COMPOSITE INDEX OF THE KENT-ROSANOFF FREE ASSOCIATION TEST

DAVID SHAKOW AND E. MORTON JELLINEK[1]
Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland

A composite index that combines the scores of most common, individual, and unusual responses is proposed for the Kent-Rosanoff Association Test. Norms, based on index scores for 200 normal Ss, as well as scores for 100 schizophrenic Ss are given.

Various studies using the Kent-Rosanoff Free Association Test (1910) have found it to be an instrument which significantly differentiates schizophrenic from normal subjects. A composite index, a special measure based on various easily derived scores from this test, has been found particularly valuable for this purpose. Although we have had many requests for a description of the index over the years, it has never been published. It seems wise to do so now, since a series of papers on the Kent-Rosanoff Free Association Test that will use this index is planned for the near future.

In scoring the Association Test, we distinguish among three kinds of response: (*a*) Most Common response—the response that has the highest frequency in the K-R tables, (*b*) Individual response—the response that is not found in the tables but that bears some relation to the stimulus word and thus meets the criteria of the "Appendix to the Frequency Tables" (Kent & Rosanoff, 1910, pp. 126–142), (*c*) Unusual response—the one that neither is found in the frequency tables nor meets the criteria of the categories listed in the appendix.

The Composite Index, devised by the authors, is derived from Paul Horst's method (1936) for obtaining a composite measure from a number of different measures of the same attribute. Horst's principle is to combine the separate measures in such a way "that the composite measure will result in giving the maximum difference between all possible pairs of members in the group [pp. 53–4]." He develops a highly complicated scheme, and the computation involved is quite intricate. Fortunately, there is a good reliable approximation to his method. It boils down to the following:

Composite Score

$$= \frac{Sr_{i1}}{s_1} X_1 + \frac{Sr_{i2}}{s_2} X_2 + \cdots\cdots + \frac{Sr_{in}}{s_n} X_n$$

where Sr_{i1}, etc., are the sums of the columns of the matrix of intercorrelations of the separate measures involved, and s_1, s_2, etc., are the standard deviations of X_1, X_2, etc.

Intercorrelations were computed on K-R variables for 200 normal control subjects examined at the Worcester State Hospital. About two thirds were in the attendant group, about one sixth in the student group, and about one sixth in the staff group. The median chronological age was about 27, and the median educational level ran between 10 and 11 years.

The following matrix of intercorrelations (see Table 1), which denotes most common responses by 1, unusual responses by 2, and individual responses by 3, is obtained. The standard deviations are: $s_1 = 10.696$, $s_2 = 8.674$, $s_3 = 3.561$. This gives us the following composite score: Composite score $= -.0051\ X_1 + 0.0608\ X_2 + .1511\ X_3$. Since this would result in a very small index, all the weights are multiplied by 10.

TABLE 1

MATRIX OF INTERCORRELATIONS

	1	2	3
1	1	−.5377	−.5269
2	−.5377	1	.0650
3	−.5269	.0650	1
Totals	−.0546	.5273	.5381

---

[1] Deceased 1963. At that time Visiting Professor of Psychiatry, Stanford University.

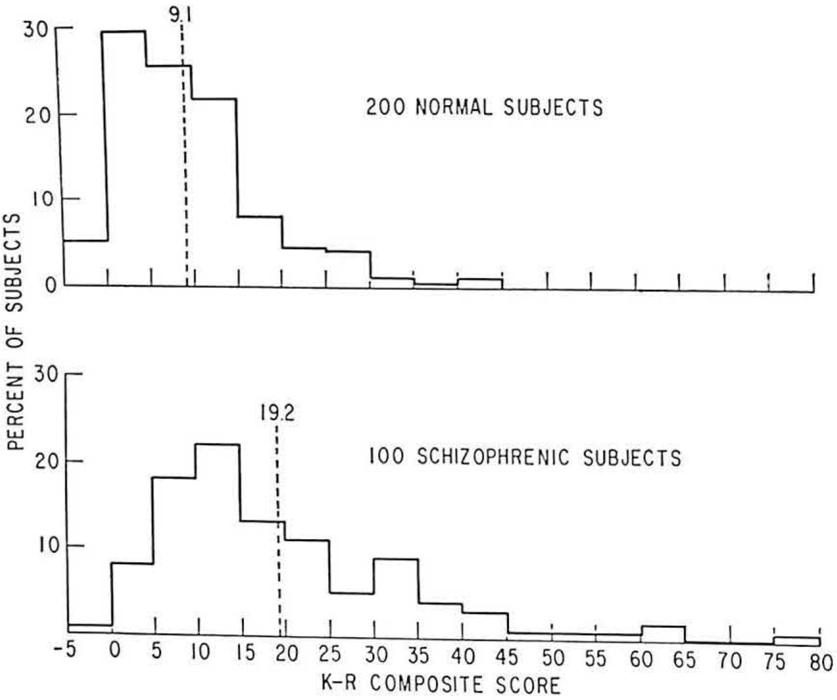


FIG. 1. Comparison of frequency distribution of 200 normal subjects and 100 schizophrenic subjects on the Kent-Rosanoff Composite Index in terms of percentage.

It will be noted that very logically the number of most common responses obtains a negative weight. The largest weight attaches to individual responses.

One hundred schizophrenic patients were also scored on the basis of the above equation. The schizophrenic group had a median age of about 31, a median educational level between eighth and ninth grade, and a median hospitalization period of 2 years.

TABLE 2

COMPOSITE INDEX FOR THE KENT-ROSANOFF ASSOCIATION TEST

	N	*M*	*SD*
Normal	200	9.1 ± 0.56	7.9
Schizophrenic	100	19.2 ± 1.47	14.7

Note.—Difference = 10.1 ± 1.57; CR = 6.4.

As shown in Table 2, the mean score of the normal subjects is 9.1 ± 0.56, with a standard deviation of 7.9. A rather skewed distribution is shown in Figure 1. The schizophrenic patients yield a mean of 19.2 ± 1.47, with a standard deviation of 14.7. Their frequency distribution is also given in Figure 1.

The difference between the two indices is 10.1 ± 1.57. The critical ratio is 6.4 ($p = < .01$), and thus the difference is very highly significant. Comparison of the frequency distributions, which are in terms of percentage, shows that they also are rather different. It will be noted that scores of five and below are found in 35% of the normal subjects, whereas only 9% of the schizophrenics give indices as low.

REFERENCES

HORST, P. Obtaining a composite measure from a number of different measures of the same attribute. *Psychometrika,* 1936, 1, 53-60.

KENT, GRACE H., & ROSANOFF, A. J. *A study of association in insanity.* Baltimore: Lord Baltimore Press, 1910.

(Received December 31, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 405–411

IMPAIRMENT OF KNOWLEDGE OF UNUSUAL MEANINGS OF FAMILIAR WORDS IN BRAIN DAMAGE AND SCHIZOPHRENIA[1]

ALLEN WILLNER
Menninger Foundation, Topeka, Kansas

A test measuring knowledge of unusual meanings of familiar words was constructed, and was hypothesized to be more sensitive than traditional vocabulary tests to language impairment in schizophrenic and brain-damaged patients. In addition, the oft-noted inferiority of these patients on tests of abstract reasoning, as compared to vocabulary tests, was hypothesized to be a function of insensitivity of usual vocabulary tests. The new test provides evidence supporting both hypotheses in schizophrenic but not brain-damaged patients. The results suggest that the similar outcome of impaired abstract reasoning in brain damage and schizophrenia may result from different processes in each disorder.

The usual vocabulary test seems primarily concerned with the number of words known, and thus emphasizes unfamiliar words used infrequently in conversation. One may, however, measure another aspect of vocabulary, that is, knowledge of the varied multiple meanings of familiar words. This variable is neglected by the usual vocabulary test which requires only that one meaning be given for each word and does not differentiate between obvious and unusual meanings. Zipf (1945) has demonstrated that familiar words have many more meanings than unfamiliar ones, and that the number of different meanings is proportional to the word's frequency of occurrence. Vocabulary tests can thus measure two dimensions: the number of words of progressively decreasing familiarity for which the subject can give one meaning, and the number of unusual meanings of familiar words known by the subject. It seems likely that standard vocabulary tests are heavily weighted with the former variable, whereas a test of the unusual meanings of familiar words is not.

A simple experiment illustrates this point. The ranked difficulty level of the vocabulary items from both the WAIS (1955) and Wechsler-Bellevue tests was correlated with the Thorndike-Lorge word-frequency count for the items. The obtained correlations for WAIS and Wechsler, −.90 and −.84, respectively ($p < .001$), indicate that difficulty level on these vocabulary tests is highly negatively associated with the familiarity of the stimulus words. A similar procedure for the stimulus words and correct answers, respectively, of the Unusual Meanings Vocabulary test (described below) resulted in nonsignificant correlation, −.13 and +.19.

A test of the unusual meanings of familiar words may thus complement the usual vocabulary test in several ways: (*a*) It measures vocabulary by assessing knowledge of the wide variety of meanings of familiar words, rather than knowledge of unfamiliar words. Empson (1951) has stressed the importance of unusual meanings of familiar words for different shadings of meaning and more articulated use of language. (*b*) It requires a degree of flexibility, since the subject must shift from the usual meaning of a word to a more unusual one. (*c*) It may be less susceptible to solution by simple associative processes since the usual meaning of the word, a very obvious synonym for example, is not the correct answer.

WILLNER UNUSUAL MEANINGS VOCABULARY TEST

The Willner Unusual Meanings Vocabulary Test (WUMV)[2] consists of 42 multiple-choice items presented on the test form in order of difficulty. The task is to select one of the five choice words which is closest in meaning to the stimulus word. The instructions specifically state, "you will find that in this vocabulary test, the correct answer will almost always be an unusual meaning of the word and not the first meaning that comes to mind." Two sample items are, *air—pen, tune, sun, lamp, burn,* and *shoot—rifle, sprout, cap, duck, door.* The stimulus words and multiple-choice alternatives are typically about five letters long, have one syllable, and are very familiar words. (The five most unfamiliar words in the entire test are: *degraded, sprout, smear, mug,* and *nude.*) The test is difficult, nevertheless, because the subject must know unusual meanings for the words. The comparative familiarity of the words in the WUMV, WAIS, and Wechsler-Bellevue I vocabulary tests can be indicated by the median word frequency level of items in each test, and the corresponding recommended school grade level for words of that frequency, using Thorndike and Lorge's (1944) criteria. The WUMV stimulus words have a median word frequency level between 99 and 100 per million words, which is considered appropriate for the third-grade level. The corresponding values for WUMV correct answers, WAIS, and Wechsler stimulus words are, respectively: between 50 and 99 per million, third-grade level; 9 per million, seventh- and eighth-grade level; and 4 per million, tenth-grade level.

---

[1] This research was supported by Grant M-4610 from the National Institute of Mental Health, United States Public Health Service.

The author is indebted to Ralph Sherfey and Bay Chotlos for their help in collecting and tabulating the data for this experiment.

[2] Copies of this test, and the WSA mentioned below, and norms for converting raw scores to

The WUMV has been standardized on a sample of 131 airmen at Forbes Air Force Base in Kansas, and norms have been compiled which permit transformation of raw scores into standard scores. The corrected split-half reliability (odd versus even items) for the WUMV is .93. The test also includes association foils, incorrect answers which have a strong association link with the stimulus word. For example, the stimulus word *shoot* in the sample item above has a strong association connection with the incorrect answer, *rifle.* Half of the 42 items in the test have association foils; these were included to penalize response in terms of automatic associative process. The corrected split-half reliability (association foil versus nonassociation foil items) for the WUMV is .92, which suggests that both parts of the test measure the same variable. The correlation between the WUMV and WAIS vocabulary tests has been determined for two samples of normal subjects: 131 airmen in the normative group and 135 university students, $r = .72$ and .64, respectively, ($p < .001$).

VOCABULARY AND ABSTRACTION

The experiments reported here investigate several hypotheses about the often-reported observation that schizophrenic and brain-damaged subjects score higher on vocabulary tests than on abstraction tests. Babcock (1930, 1933), and also Shipley (1940) explained this finding by asserting that vocabulary is less vulnerable to impairment in schizophrenic and brain-damaged subjects because it is learned earlier and hence retained better than abstract reasoning. Yacorzynski (1941) criticized Babcock's hypothesis, and explained the finding of higher scores on vocabulary tests on the basis that vocabulary tests are too easy, that is, that they permit several paths to a correct answer. He cited two types of imprecision on vocabulary tests: (*a*) Any approximate answer is acceptable on these tests, whether it is abstract, concrete or even somewhat peripherally relevant. There have been several attempts to remedy this defect since 1941 (cf. Feifel, 1949) and a number of contemporary vocabulary tests such as the WAIS Vocabulary test (Wechsler, 1955) do make some provision for it. (*b*) Yacorzynski also noted that most words have several meanings, and that a person with intellectual impairment resulting from brain damage or schizophrenia might lose all the meanings of a word except the easiest one, and yet not be penalized in his score, since any meaning is acceptable. He recommended that a vocabulary test be constructed which would be sensitive to the loss of unusual meanings of words. Yacorzynski's second criticism seems related to observations by Bleuler (1911) and Vigotsky (1934). Bleuler stressed the primary role of association dis-

---

standard scores may be obtained from Allen Willner, Menninger Foundation, Box 829, Topeka, Kansas.

order in schizophrenia and noted the frequent disruption of word meanings:

> even our commonly used and presumably simple concepts change with the context. For example, the concept water varies greatly with its context, as in chemistry, physiology, shipping, landscape, flood, power supply and so on. Every one of these special concepts is tied to our ideas by a different thread; no healthy person thinks of soda water when a flood is sweeping his house away, nor of buoyancy when he wants to quench his thirst [p. 588].

Vigotsky (1934) also focused on changes in the meanings of words as the basic element in schizophrenic impairment of abstract reasoning:

> Whatever may be its cause and paradoxical as it may appear, the fact is nevertheless fairly well established that the meanings of words become pathologically altered in schizophrenia, though such alterations do not become apparent for a long time [p. 1071].

Yacorzynski's recommendation that a vocabulary test be constructed which is sensitive to loss of the unusual meanings of words has been carried out by several investigators, including Capps (1939), Acklesberg (1944), and the present writer. Capps, using a "homograph" test asked his subjects to give as many meanings as they could to several words ranging from familiar to quite unfamiliar ones. He found that severely brain-damaged subjects gave fewer meanings than mildly or moderately brain-damaged subjects. However, Capps also found that his brain-damaged subjects performed just as poorly when asked only to say as many words as came to mind in 2 minutes. Finally, he found that his subjects performed at the same level on a traditional vocabulary test, the Terman Vocabulary Test. The subjects thus showed no significant differences in performance on any of the three vocabulary tests. These results were later replicated by Acklesberg (1944) using Capps' tests.

HYPOTHESES

The three experiments reported here investigate the following hypotheses:

1. The WUMV can detect the disruption of word meanings in schizophrenia hypothesized by Bleuler, Vigotsky, and Yacorzynski.
2. Schizophrenic and brain-damaged patients typically obtain higher scores on vocabulary tests than tests of abstraction because the ordinary vocabulary test is rather insensitive to losses in language functioning (Yacorzynski's point). The first experiment investigates Hypothesis 1; the second and third experiments investigate Hypothesis 2 with schizophrenic and brain-damaged subjects, respectively.

EXPERIMENT I

Subjects

This experiment compared 51 pairs of schizophrenic and control patients from the Topeka, Kansas Veterans Administration Hospital. The former were chronic schizophrenic patients, and the latter general medical, psychosomatic, and psychoneurotic patients who were neither psychotic, brain damaged, nor alcoholic. Each pair was exactly matched in WAIS vocabulary weighted score and very closely matched in age and education. (The schizophrenic subjects were nonsignificantly higher in age, by .21 years, and in education, by .31 years.)

Results

The control group scored significantly higher ($p < .001$) on the WUMV, strongly suggesting that the schizophrenics show impaired knowledge of unusual meanings of familiar words. The mean score for control and schizophrenic groups, respectively, was 22.04 and 17.53; the control group mean score was thus 26% higher.

TEST BATTERY

Experiments II and III both use the same three tests: WUMV, WAIS vocabulary, and Willner-Scheerer Analogy test (WSA). The WSA is a new analogy test, constructed so as to be unsolvable by word association. The WSA is described in a monograph, Willner (1964), reporting an investigation which demonstrated that approximately half the analogy items on current tests can be solved by automatic word association rather than abstract conceptualization.

The three-test battery was administered to 131 airmen at Forbes Air Force Base in Kansas, to provide normative data and standard scores for all three tests. (The raw scores for each test are converted to standard scores with a mean of 10 and standard deviation of three—the same formula used for

TABLE 1

ANALYSIS OF VARIANCE FOR SCORES OF SCHIZOPHRENIC SUBJECTS ON THREE TESTS

Source	*SS*	*df*	*F*
Uncorrelated			
Order	50.04	1	1.88
Between *S*s in same group	2,495.70	94	
Correlated			
Tests	241.19	2	38.28**
Order × tests	1.40	2	0.22
Pooled *S*s × tests	592.08	188	
Total	3,380.41	287	

Means

	Scheffé's test		
	$\bar{X}_1$	$\bar{X}_2$	t
WAIS versus WUMV	10.55	8.76	7.16*
WUMV versus WSA	8.76	8.49	1.08
WAIS versus WSA	10.55	8.49	8.24*

* $p < .005$.
** $p < .001$.

WAIS scaled scores.) This procedure permits direct comparison of scores on the three tests.

EXPERIMENT II

Subjects

The subjects were 96 chronic schizophrenic patients from the Topeka Veterans Administration Hospital (none of whom were subjects in Experiment I), having a mean age of 38.28 years and mean educational level of 11.52 years. The diagnoses of the subjects were distributed as follows: paranoid 52, undifferentiated 28, catatonic 11, simple 3, schizoaffective 2.

Procedure

All subjects took the WAIS vocabulary, WUMV, and WSA, with the order of presentation of the two vocabulary tests counterbalanced so that each was presented first one-half the time, to control for order effects such as practice and fatigue.

Predictions

1. The subjects will score significantly higher on the WAIS vocabulary than the WSA. This prediction is in accordance with the typical finding that schizophrenic subjects score higher on standard vocabulary than on abstraction tests.

2. The subjects will score significantly higher on the WAIS than on the WUMV. This is predicted because the WUMV is hypothesized to be more sensitive to language impairment in schizophrenia than the WAIS vocabulary.

3. There will be no significant difference in subjects' scores on the WUMV and WSA. This prediction follows from the hypothesis that the typical finding of better performance by schizophrenics on vocabulary than abstraction tests occurs because the vocabulary test is not sensitive enough to language impairment in the schizophrenic. Since the WUMV is hypothesized to be sensitive to such impairment, no difference in WUMV and WSA scores is expected.

Results

An analysis of variance, Table 1, shows that order of presentation of the two vocabulary tests has no significant effect on the subjects' scores. The variance attributable to differential performance of the subjects on the three tests, however, is highly significant ($p < .001$). A series of individual mean comparisons of the subjects' scores on the three tests, using Scheffé's test (see Table 1) provides evidence concerning the three predictions: (*a*) The mean score for the WAIS was significantly higher ($p < .005$) than for the WSA, confirming the first prediction. The subjects scored .70 standard deviation units higher on the WAIS. (*b*) The mean score for the WAIS was significantly higher than the WUMV ($p < .005$), confirming the second prediction. The subjects scored .60 standard deviation units higher on the WAIS. (*c*) The mean score for the WUMV was not significantly different from the WSA, confirming the third prediction. The subjects scored only .09 standard deviation units higher on the WUMV. The results thus confirm all three predictions and support the hypothesis that the minimal impairment evidenced by schizophrenics on usual vocabulary tests is a function of the insensitivity of these tests. Schizophrenic subjects demonstrate as much impairment on the WUMV as on the analogy test.

Since predictions one and two were confirmed the data were next analyzed to see the percentage of the schizophrenic subjects correctly identified, that is, having a WAIS vocabulary score higher than the WSA, or higher than the WUMV. The former comparison correctly identified 71% of the subjects, and the latter also identified 71% correctly. When subjects obtaining the *same* score, on both the WAIS and the other test compared with it, are omitted, the percentage of correct classification rises to 82 for the WSA and 79 for the WUMV.

One may also question whether the WUMV is composed of two classes of items, that is, items testing only the unusual meaning function and items which in addition have association foils and may test substantially different variables. Since half the test items have association foils, and half do not, it is possible to correlate these scores and obtain a corrected split-half reliability correlation coefficient. The correlation is .92 for the normative group and .90 for the schizophrenic subjects, strongly suggesting that association-foil and nonassociation-foil items measure the same variable in both normal and schizophrenic subjects. Although association-foil (AF) and nonassociation-foil (NAF) items were so highly correlated, it was still possible that AF items were more difficult for subjects generally, and especially so for the schizophrenic subjects. The AF items were significantly more difficult than the NAF items for both the normative group ($p < .001$) and the schizophrenic group ($p < .001$). There was no significant difference, however, between normative and schizophrenic groups in the differential between mean score on AF and NAF items. Both types of subjects score more poorly on AF items, and to about the same degree.

EXPERIMENT III

Subjects

The subjects were 36 brain-damaged patients from the Topeka Veterans Administration Hospital with a mean age of 47.2 years and mean educational level of 11.10 years. The distribution of diagnostic categories for subjects were as follows: chronic brain syndrome 24, acute brain syndrome 2, convulsive disorder 5, posttraumatic encephalopathy 2, subdural hematoma 1, postencephalitic Parkinsonism 1, Parkinsonism syndrome 1.

Procedure

This experiment is essentially a replication of Experiment II with brain-damaged subjects, to test the generality of the findings. All subjects took the WAIS vocabulary, WUMV, and WSA tests.

Predictions

The predictions are the same as in the preceding experiment. The subjects will score significantly higher on the WAIS vocabulary than on the WSA. The subjects will score significantly higher on the WAIS than on the WUMV. There will be no significant difference in subjects' scores on the WUMV and the WSA.

TABLE 2

ANALYSIS OF VARIANCE FOR SCORES OF BRAIN-DAMAGED SUBJECTS ON THREE TESTS

Source	*SS*	*df*	*F*
Subjects (A)	988.7	35	7.15**
Tests (B)	128.7	2	16.30**
A × B	276.6	70	
Total	1,394.0	107	

Means

	Scheffé's test		
	$\bar{X}_1$	$\bar{X}_2$	*t*
WAIS versus WUMV	9.81	9.00	1.72
WUMV versus WSA	9.00	7.19	3.85*
WAIS versus WSA	9.81	7.19	5.57*

* $p < .005$.
** $p < .001$.

Results

An analysis of variance, Table 2, shows that the variance attributable to differential performance of the subjects on the three tests is highly significant ($p < .001$). A series of individual mean comparisons of the subjects' scores on the three tests, using Scheffé's test, Table 2, provides evidence concerning the three predictions: (*a*) The mean score on the WAIS was significantly higher than the WSA ($p < .005$), confirming the first prediction. The brain-damaged subjects scored .87 standard deviations higher on the WAIS than the WSA. (*b*) The mean score on the WAIS was not significantly higher than the WUMV, so that the second prediction was not confirmed. The results were in the predicted direction, however, with the subjects scoring .27 standard deviations higher on the WAIS. Inspection of the means for schizophrenic and brain-damaged subjects shows that the latter scored considerably lower than the schizophrenics on the WAIS vocabulary, but slightly better than them on the WUMV. (*c*) The mean score for the WUMV was significantly higher than the WSA ($p < .005$), so that the third prediction was not confirmed. The subjects scored .60 standard deviations higher on the WUMV.

The WUMV scores for these brain-damaged subjects were analyzed to see whether the AF and NAF items measured the same variable. The corrected split-half reliability coefficient (AF versus NAF items) was .94. The corresponding correlations of .92 for the normative subjects and .90 for schizophrenic subjects strongly suggest that AF and NAF items measure the same variable for all three types of subjects. The brain-damaged subjects obtained significantly lower scores on the AF items ($p < .001$) as did the other two groups. The differential between mean score on AF and NAF items, however, was significantly larger for the brain-damaged subject than either the normative subjects ($p < .001$) or the schizophrenic subjects ($p < .01$). These results suggest with regard to AF and NAF items on the WUMV: They measure the same variable for all three groups. AF items are harder than NAF items for all three groups. Brain-damaged subjects find AF items significantly harder than do either normal or schizophrenic groups.

DISCUSSION

The data provide strong support for all hypotheses and predictions concerning schizophrenic subjects. For the brain-damaged subjects, the first prediction was confirmed and Predictions 2 and 3 were not. The nonconfirmation of Prediction 2 is consistent with results reported by Capps (1939) and Acklesberg (1944). They found general loss on all vocabulary tests, with no significant differential loss between a standard vocabulary test and a homograph test which required that the subject give several meanings for words. The nonconfirmation of Prediction 3 is consistent with the results of Capps (1939) who found that brain-damaged subjects performed more poorly on a test of abstract reasoning, a categorizing test, than on any of the vocabulary tests used, including a homograph test and a standard vocabulary test. Capps and Acklesberg, however, did not use a homograph test or other test of knowledge of unusual meanings with schizophrenic subjects. There has been no experiment published, to the writer's knowledge, where performance on standard vocabulary tests, tests of unusual meanings of words, and tests of abstraction has been studied in both brain-damaged and schizophrenic subjects. The results of the present experiment strongly suggest that, although both brain-damaged and schizophrenic subjects show severe loss on a test of abstraction, the WSA, only the schizophrenic subjects show severe loss on a test of the unusual meanings of words, the WUMV.

These findings also suggest that tests such as the WAIS vocabulary, WSA and WUMV might be of some aid in differentiating brain-damaged from schizophrenic patients. Both groups might be expected to score higher on the WAIS than the WSA. Only the schizophrenic subjects would be expected to score higher on the WAIS than the WUMV, while only the brain-damaged subjects would be expected to score higher on the WUMV than the WSA.

A speculative interpretation of the findings is that abstract reasoning in schizophrenia and brain damage may be impaired as a consequence of two different processes:

Impairment of abstract reasoning in schizophrenia may be a function of *pathological alterations in the meanings of words used* in the abstraction process. Test findings of impaired performance of schizophrenic subjects on both an analogy test (WSA) and a test of the unusual meanings of familiar words (WUMV) are consistent with this hypothesis. Bleuler (1911) and Vigotsky (1934) have expressed similar views, that is, that impairment of abstract reasoning in schizophrenia is a function of an association disturbance which pathologically alters word meanings.

Impairment of abstract reasoning in brain damage may be a function of *diminished ability to form relationships between words,* and *not* a consequence of pathological alterations in the meanings of words. Test-findings of impaired performance of brain-damaged subjects on an analogy test (WSA), but not on a test of the unusual meanings of familiar words (WUMV), are consistent with this hypothesis. Capps' and Acklesberg's findings of severe impairment of abstract reasoning in brain-damaged patients, with much smaller losses on vocabulary tests, and no differential losses in number of different meanings of words known, are also consistent with this hypothesis.

After formulating this speculative interpretation of the results, the writer found that Schilder (1930) had advocated a similar explanation. Schilder contrasted thought disorder in schizophrenia and general paresis:

> The schizophrenic's thought disorder encroaches on the concept 'in statu nascendi,' while that of a paretic attacks a relatively developed concept. Each involves, so to speak, a different phase of the thought process. In other words: the schizophrenic's thought disorder affects the core, the paretic's the periphery of experience. . . . Schizophrenic thought disorder affects the development of concepts and propositions, the paretic affects the developed concepts and propositions [p. 563].

(Here by *concepts* Schilder refers to the cluster of different meanings and associations surrounding common words and ideas.) "A comparison of the concepts of our paretics with those of schizophrenics makes it obvious that the former do not differ very much from those of everyday. The meaning of officer, pope, shark to the paretic is not very different from the normal [p. 563]." Thus Schilder states that the schizophrenic thought disorder affects a very early phase of the abstraction process, by strongly distorting the word meanings, that is, concepts. The organic thought disorder is seen as a disruption of a later phase of the process, the connections between "developed concepts." The interpretation of the present experiments thus seems quite close to Schilder's conceptions, formulated three decades earlier.

REFERENCES

ACKLESBERG, S. Vocabulary and mental deterioration in senile dementia. *Journal of Abnormal and Social Psychology,* 1944, 39, 393–406.

BABCOCK, H. An experiment in the measurement of mental deterioration. *Archives of Psychology,* 1930, No. 117, 105.

BABCOCK, H. *Dementia praecox: A psychological study.* Lancaster, Pa.: Science Press, 1933.

BLEULER, E. *Dementia praecox oder Gruppe der Schizophrenien.* (Orig. publ. 1911) [*Dementia praecox or the group of schizophrenias.*] (Trans. by J. Zinkin) New York: International University, 1950.

CAPPS, H. M. Vocabulary changes in mental deterioration. *Archives of Psychology,* 1939, 34, No. 242.

EMPSON, W. *The structure of complex words.* Norfolk, Conn.: New Directions, 1951.

FEIFEL, H. Qualitative changes in the vocabulary responses of normals and abnormals. *Genetic Psychology Monographs,* 1949, 39, 159–204.

SCHILDER, P. Studien zur Psychologie und Symptomatologie der progressiven Paralyse. (Orig. publ. 1930) In D. Rapaport, *Organization and pathology of thought.* (Trans. by D. Rapaport) New York: Columbia University Press, 1951.

SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology,* 1940, 9, 371–377.

THORNDIKE, E. L., & LORGE, I. *The teachers word-book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

VIGOTSKY, L. S. Thought in schizophrenia. *Archives of Neurology and Psychiatry,* 1934, 81, 1063–1077.

WECHSLER, D. *Wechsler-Bellevue Intelligence Scale.* New York: Psychological Corporation, 1939.

WECHSLER, D. *Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

WILLNER, A. An experimental analysis of analogical reasoning. *Psychological Reports Monograph Supplement,* 1964, 15(3), 479–494.

YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock deterioration test. *Psychology Review,* 1941, 48, 261–267.

ZIPF, G. K. The meaning-frequency relationship of words. *Journal of Genetic Psychology,* 1945, 33, 251–256.

(Received January 11, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 412–418

DIFFERENTIAL EFFECTS OF SOCIAL REWARD AND PUNISHMENT ON DEPENDENT AND DEPENDENCY-ANXIOUS SCHIZOPHRENICS[1]

ARNOLD R. GOLDMAN[2]

Veterans Administration Hospital, Fort Meade, South Dakota[3]

Several recent studies have reported mild social punishment to be more effective than social reward for motivating schizophrenics' learning. However, theoretical formulations of schizophrenia, as well as learning studies with normals, suggest that the relative effectiveness of positive and negative reinforcement varies with the individual's dependency needs and conflicts. This study was therefore designed to determine whether social rewards and social punishments would have differential effects upon the learning performance of highly dependent (HD) and highly dependency-anxious (HD-A) schizophrenics. In accord with theoretical expectations, the dependent Ss performed better under rewards than under punishment, while the dependency-anxious Ss performed better under punishment than under rewards. However, these differential reactions extinguished within a surprisingly short period.

In several recent studies, the performance of schizophrenics was found to improve rapidly under conditions of mild social punishment, but had failed to improve or had even declined under conditions of mild social reward (Atkinson & Robinson, 1961; Karras, 1962). Since normals usually respond better to rewards than to punishments, it has been presumed that the obverse reactions by schizophrenics reflect a distrust of rewards, a low readiness to perform, or a hypersensitivity to censure. To date, however, little or no effort has been made to test these "motivational" interpretations directly or in depth. Indeed, the research designs in the vast majority of these learning investigations have been categorical or nominal. Groups of schizophrenics tend to be selected without regard to personality or motivational considerations and compared with one another or with groups of normals having equally distinct and varied personalities. It is therefore somewhat understandable that on several occasions both punishments and rewards have failed to produce the expected effects (D'alessio & Spence, 1963; Fischer, 1963).

Indeed, studies with normals have indicated for some time that both rewards and punishments can enhance as well as impair performance, depending, for example, upon the subject's manifest anxiety (MA) level (Taylor, 1956), his achievement needs (McCelland, Atkinson, Clark, & Lowell, 1953), and his dependency needs (Bernadin & Jessor, 1957). All things considered, it would seem appropriate that the effects of social reinforcers on schizophrenics' performance also be tested in light of specific personality attributes. It was thought that the most relevant personality dimension might be dependency-anxiety, since dependency conflicts have long been considered basic in the etiology and treatment of schizophrenic disorders (Fromm-Reichmann, 1959). The purpose of this study, therefore, was to employ a dimensional rather than categorical design, whereby positive and negative social reinforcements could be studied in terms of their effects upon dependent and dependency-anxious schizophrenics.

---

[1] Presented in part at the annual meeting of the American Psychological Association, Philadelphia, 1963, and based upon a doctoral dissertation submitted to the Department of Psychology, Stanford University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The investigation was carried out while the author was a research trainee at the Veterans Administration Hospital, Palo Alto, California.

[2] The author wishes to express his gratitude to the members of his dissertation committee: Albert Bandura (Chairman), Gregory Bateson, Ernest R. Hilgard, William M. McCord, and Paul W. McReynolds. Considerable thanks also are due to Quinn McNemar, David L. Cressler, William C. Horne, Marvin Sherman, and William T. Query for their invaluable aid and assistance.

[3] Now at the Rehabilitation Center, Philadelphia State Hospital.

According to some learning theorists (e.g., Bandura & Walters, 1959), the dependency motive, with its concomitant need for approval, affection, and attention, develops early in childhood through association with primary gratification. Eventually, the socialization process intervenes, and the child's attention- and approval-seeking behaviors undergo some modification. However, should the child's dependency bids meet with very strong parental rejection or with severe punishment, the child learns accordingly to avoid any expressions of dependency. In extreme cases, as may be reflected in schizophrenia, the individual may manifest suspicion, fear, and strong resistance to entering practically all dependent or affectional relationships. For these persons, praise, attention, and approval by others can not be properly labeled as social rewards so much as social punishments, for they serve principally as cues which arouse his anxiety, distrust, and discomfort.

Because these reactions are generally incompatible with the demands of an experimental learning situation, the dependency-anxious subjects would not be expected to perform as well as the dependent subjects under conditions of social reward. To demonstrate this effect, Cairns (1959) administered mild praise and approval to high and low dependency-anxious delinquents during the course of their performance in a verbal conditioning and in a paired-associate learning situation. As predicted, the low dependency-anxious delinquents showed progressive improvement, whereas the high dependency-anxious delinquents demonstrated progressive impairment.

Similarly, in the present study, it was predicted that under conditions of social reward, the learning performance of schizophrenics having strong dependency needs (with little, if any, accompanying anxiety) would be superior to that of patients who feared dependency involvements. Presumably, the former would strive to perform effectively in order to sustain the experimenter's praise and approval, while the latter would be made anxious by the experimenter's "nurturance" or reward. However, in a somewhat hostile or at least non-nurturant atmosphere, that is, where only social punishments were applied, high dependency-anxious (HD-A) subjects should experience relatively little anxiety. Consequently, HD-A subjects would perform more effectively under punishment than they had under conditions of social reward. Moreover, it should follow that under social punishment conditions, HD-A subjects would perform more adequately than highly dependent (HD) subjects, since punishment would be anxiety-arousing for the latter and tend to disrupt their performance. Finally, under conditions of nonevaluation, HD and HD-A subjects should do equally well, since neither group would experience undue interference from anxiety-arousing stimuli.

METHOD

Selection of Dependent and Dependency-Anxious Subjects

Since few guidelines exist for the systematic assessment of dependency-anxiety in institutionalized adult schizophrenics, new measures were developed for the purpose of distinguishing the dependency-anxious from the highly dependent patients. It was assumed that dependency-anxiety would be indicated in patients who persistently resisted offers of help, assistance, or attention, despite the fact that the instigation for dependency was strong. Conversely, if a patient consistently sought the help and support of others for problems which he could very well handle on his own, an interpretation of high dependency would then seem justified. The following overt and thematic measures were therefore devised and used jointly in selecting the patient sample:

Ward behavior classifications. A total of 11 ward administrators (8 psychiatrists and 3 clinical psychologists) classified each of the schizophrenics on their ward as either "approachers," "avoiders," or "neutrals," depending upon whether the patient typically entered into or typically avoided dependency involvements. To be classified as an approacher (i.e., as an HD subject), a patient was to have frequently sought the administrator's attention and approval, or repeatedly and unnecessarily asked for his advice, assistance, help, or consolation. Avoiders (HD-A subjects) were defined essentially as patients who consistently had shown themselves strongly reluctant to ask for their administrator's help, advice, or assistance, or who had rejected such offers, especially when faced with obvious difficulties.

As a check on the stability of these classifications, the administrators rerated their patients after an interval of approximately 1 month. The median percentage of intrajudge agreement was a very satisfactory 94.

Incomplete story test. A series of 15 incomplete stories was administered to the subjects classified as approachers or avoiders. The stories had been drawn from a larger pool on the basis of their effectiveness in differentiating pilot subjects. Eight stories portrayed help-seeking situations, three involved attention-seeking, and three approval-seeking situations. One neutral story was used to introduce the task. The protagonist in each story was depicted as initiating a dependency bid, and subjects were asked to complete the story by anticipating the response of others to his dependency request. To heighten interest and involvement, each subject was asked to imagine himself as the protagonist, and to describe the expected outcome. The number of stories in which the patient indicated that his hypothetical bid for dependency was rejected or punished was summed to yield a total dependency-anxiety score. Based on the pretest median, the subjects who anticipated rejection in five or more stories were characterized as HD-A, and those with scores of less than five were designated as HD and relatively free of dependency-anxiety. A comparison of the author's ratings with those of an advanced graduate student resulted in a very satisfactory estimate of scoring reliability ($r_t = .99$).

The situational test. After completing the thematic test, subjects were presented with a near-insoluble form-board puzzle known as the Kwazy Quilt. The task required the subjects to return 13 plastic pieces to their original position on a form board (see Cairns & Lewis, 1962, for further details). The examiner, seated alongside the subject, introduced the task as a very difficult puzzle, "and so if you would want some help with it, just ask me, and the two of us can work on it together . . . ." The time which elapsed before the subject made a direct and explicit request for help was regarded as a measure of his resistance to dependency involvement. If no request was made, the session was terminated at the end of 20 minutes, during which time the subject would have received two additional offers of assistance. Scores above 9 minutes (pretest median) were considered to indicate the presence of dependency-anxiety, and those below as reflecting high dependency with relatively little accompanying anxiety.

Subjects

Of the 416 patients available for classification, 56% (233) were categorized as avoiders and 32% (133) as approachers. The 50 patients who received a neutral classification were eliminated from further study, as were 33 acutely disturbed patients. Another 83 patients were unavailable for testing, while 60 subjects were held in reserve, leaving 190 patients to be tested on the thematic and situational tests. From this group, 89 patients were eliminated because of marked discrepancies in performance on the screening measures. From the patients remaining, a group of 36 HD subjects and 36 HD-A subjects were selected on the basis of their consistently HD or consistently HD-A scores.[4]

These two groups proved to be almost identical with regard to age (means of 37), years of education (11.5), race, and marital status. However, the HD-A schizophrenics were the more chronic group, having had a significantly longer period of hospitalization (79 months, compared to 56, $t = 2.69$, $p < .02$), a slightly greater number of hospitalizations ($p < .30$), and slightly higher severity of illness ratings ($p < .30$) than had the HD schizophrenics.

TABLE 1

PAIRED-ASSOCIATE LISTS

List I[a]	List II	List III	List IV
SHORT—1	RAPID—6	HARD—7	CALM—16
FRAIL—2	QUICK—8	STURDY—9	HUSHED—17
SMALL—3	FAST—10	SOLID—11	QUIET—18
TINY—4	SWIFT—12	FIRM—13	SILENT—19
LITTLE—5	SPEED—14	STRONG—15	STILL—20

[a] Presented to all subjects under conditions of nonevaluation.

The Learning Task

Many of the words comprising the standard paired-associate lists proved to be so unfamiliar to the pilot subjects that they seldom attempted to guess the correct response word. Since social reinforcements could not be administered appropriately unless the subjects verbally anticipated each response word, new lists were developed in which the stimulus words (synonyms) were paired with more readily familiar Arabic numbers. The four lists were found to be of equal difficulty and were presented to the 72 experimental subjects in the order of their appearance in Table 1.

Testing was done individually, with the subject seated in front of an electrically driven Hull-type memory drum. Each stimulus word appeared independently in the aperture for 2.5 seconds, after which both the word and its associate number were exposed together for an additional 2.5 seconds. Each trial was therefore completed in exactly 25 seconds, with a 7.5-seconds interval maintained between trials. The first time through each list, the five word pairs appeared in the aperture, and the subject was required to say them aloud. In subsequent trials the stimulus word appeared alone, and the subject was required to anticipate its number within 2.5 seconds, that is, before the number appeared. Hence,

---

[4] This selection yielded two rather distinct groups inasmuch as there was virtually no overlap between the subjects' ward behavior classifications, and none between the two performance indicies for 31 HD and 23 HD-A subjects. The overlap in the remaining cases was minor in the sense that these 18 patients had either scored near the median on the thematic or on the puzzle task, but not on both.

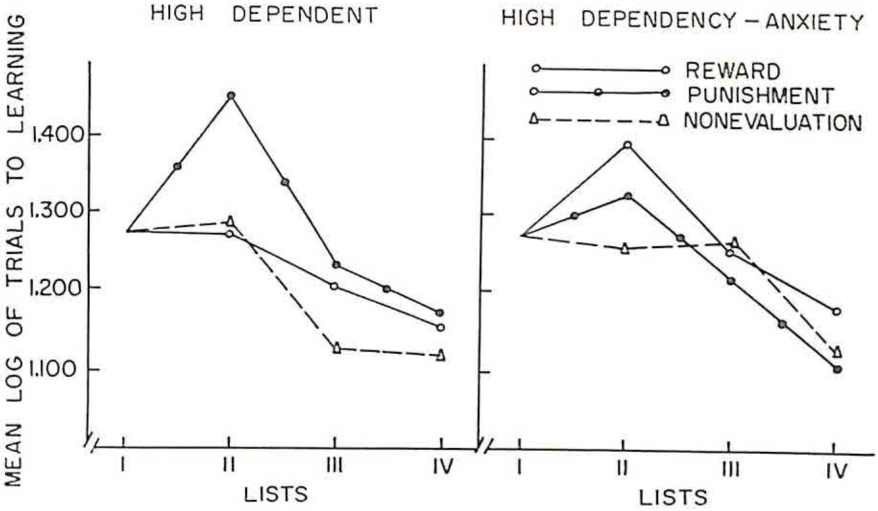


Fig. 1. Course of learning for high dependent and high dependency-anxious schizophrenics under three conditions of social reinforcement. (Scores were corrected by covariance analysis, using List I as the covariate.)

the subject could quickly determine for himself whether or not he had given a correct response.

To eliminate serial learning, the order of the word pairs was varied on successive trials in one of nine different ways. The subject's score on each list was the number of trials he required to meet the learning criterion of two errorless recitations. List I was presented to all subjects without reinforcement, in order to establish a baseline level of learning from which the effects of reinforcement could be determined.

Treatment Conditions

Six independent groups of 12 subjects were formed by randomly assigning one-third of the HD and one-third of the HD-A subjects to a social reward condition and one-third to mild social punishment; the remaining one-third served as controls and therefore performed under nonevaluation. Under the reward condition, the experimenter[5] expressed approval by saying warmly, "Right!" "Good!" "Fine!" or "Correct!" after each subject's correct anticipation, but said nothing for incorrect responses. In the punishment situation, the experimenter did not respond to correct anticipations but expressed disapproval after each error by saying "Wrong!" "No!" "No good!" or "Incorrect!"

Results

The course of learning among the six groups was analyzed in two parts, one relating to the introduction of social reinforcements on List II, and a second dealing with the amount of improvement shown between Lists I and IV. For both considerations, the learning scores (trials to criterion) were transformed to logarithms and treated by analyses of covariance, with List I as the covariate. The adjusted mean logs are presented in Figure 1 and show graphically the learning profiles of the HD and HD-A groups under the three treatment (reinforcement) conditions.

As the curves would indicate, the change from conditions of nonevaluation on List I to either reward or punishment on List II was a relatively difficult transition for subjects to make. However, the results of a 2 × 3 analysis of covariance upon patients' performance on List II (Table 2) indicated a probable main effect for reinforcements. Inspection of the three treatment means suggested that while the control subjects improved as a

TABLE 2

Analysis of Covariance on Logs of Trials Taken to Learn List II

Source	*df*	Adjusted *SS*	*MS*	*F*
Groups (A)	1	.001	.001	—
Reinforcement (B)	2	.179	.089	1.99*
A × B	2	.171	.085	1.90*
Within	65	2.928	.045	

* $.05 < p < .10$.

---

[5] Only the author administered the learning task. Subsequent analyses indicated, however, that the performance of patients who had encountered the author in both the screening and learning tasks was no different from that of subject screened by an advanced graduate student.

TABLE 3

ADJUSTED MEAN LOGS OF TRIALS TO LEARNING FOR THE INDIVIDUAL TREATMENT CELLS AND THEIR TOTALS

Groups	Introduction of reinforcement (List II[a])			Overall improvement (List IV[a])		
	Reward	Punishment	Nonevaluation	Reward	Punishment	Nonevaluation
HD	1.28	1.45	1.28	1.15	1.17	1.12
HD-A	1.39	1.33	1.26	1.18	1.11	1.13
All subjects	1.34	1.39	1.27	1.17	1.14	1.12

[a] List I as the Covariate.

group, subjects who had begun to receive either rewards or punishments on List II did not improve but showed instead a considerable decrement in learning. An analysis of the individually adjusted treatment means indicated that the performance of control subjects on List II was not reliably superior to that of the rewarded groups ($p < .10$) but was quite superior to the performance of the punished groups ($t = 2.15$, $p < .05$). Thus, without yet considering the personality of the groups, social punishment, upon its introduction, appeared to be the least effective and, in fact, the most disruptive of the three conditions.

However, turning to personality considerations, the results for the Reinforcement × Group interaction (Table 2) suggested that the relative effectiveness of reward, punishment, and nonevaluation had varied between the two groups. For a more direct test of the major hypotheses, the HD and HD-A groups' performance on List II was then evaluated under reward and punishment. This 2×2 analysis of covariance clearly demonstrated the expected interaction between groups and treatments ($F = 4.09$, $df = 1, 43$, $p < .025$). In other words, social punishment produced more impairment among the dependent than among the dependency-anxious subjects, with social rewards having the opposite effects. Indeed, under reward, it was the HD-A subjects who showed a decided decrement in performance and the HD subjects who maintained a relatively even rate of learning. The differences between the individual group means (Table 3), however, were not as sharply divided as had been expected. HD subjects tended to perform less adequately than the HD-A subjects under punishment ($t = 1.44$, $p < .08$), and more adequately than the HD-A subjects under reward ($t = 1.39$, $p < .09$). However, the discrepancy between the performance of the HD subjects under reward and under punishment was clearly significant ($t = 2.04$, $p < .02$); for the HD-A subjects, this difference was in the expected direction, but inconsequential.

General support also was provided for the predictions regarding the performance of the control subjects on List II. Thus, no significant discrepancy was observed between the HD subjects' and HD-A subjects' rates of learning under nonevaluation. Moreover, the HD subjects in the control situation did not perform any better than the HD subjects under reward, but did perform significantly better than their counterparts under punishment ($t = 1.98$, $p < .03$). For the HD-A subjects, the opposite was again the case. The performance of the HD-A subjects under nonevaluation was superior to that of the HD-A subjects under reward ($t = 1.51$, $p < .07$), but was not different under conditions of punishment.

By List IV the group differences noted above were no longer in evidence (Table 3). Indeed, the relatively consistent pattern of improvement shown across Lists II, III, and IV was found, however, to be the same for all six groups. Perhaps the only noteworthy event with regard to overall improvement was the change in the relative order of rewards, punishment, and nonreinforcement. In contrast to having been initially the most debilitating of the three conditions, punishment yielded more improvement between Lists II and IV than had rewards or nonevaluation. However, no clearly significant interactions could be observed between stages (lists) and reinforcement, or between stages and groups. This relatively uniform rate of improvement among groups would therefore mean that punishment had quickly lost whatever debilitating effects it had held initially. Indeed, punishment subsequently served to facilitate the performance of both HD subjects and HD-A subjects about as adequately as, and possibly even more adequately than, rewards and nonreinforcement.

DISCUSSION AND CONCLUSIONS

The findings differed in general from those reported by most other investigators in that social punishments had been neither necessarily more effective nor necessarily less effective than social rewards. It was found, instead, that social reinforcement effectiveness was contingent upon the patients' particular dependency needs and conflicts. Thus, dependent subjects performed better under rewards than under punishment, while dependency-anxious subjects performed better under punishment than under rewards. However, this finding was limited to the events surrounding initial learning, since no differences were observed among the groups' overall rates of improvement. Indeed, the HD and HD-A groups differed only in the amount of impairment they manifested in response to the introduction of rewards and to the introduction of punishments. Why rewards and punishments should have initially impaired both the HD and HD-A subjects, albeit differently, and why these differential effects should have extinguished so rapidly, can be explained in part by re-examining the particular learning task used and the procedures employed.

It can be noted that only the lists had changed for the subjects in the nonevaluative condition. The experimental subjects, however, also encountered a change in the role of their experimenter, from nonevaluator on List I to rewarder or punisher on List II. The experimenter's sudden and active participation on List II was probably confusing and distracting to the patients, and interfered with their performance. This explanation becomes more plausible when it is recalled that subjects not only had but 2.5 seconds to prepare for the next response word, but had no "need" of the experimenter's evaluative comments, since they saw for themselves whether their anticipations had been correct or incorrect. Thus, in this learning task requiring both speed and concentration, the abrupt insertion of cues having no informational value probably served to disconcert the subjects. Nevertheless, these cues proved to be less disturbing for the HD subjects under rewards than for the HD-A subjects under punishment.

An "information" hypothesis (e.g., Egger & Miller, 1963) might also be applied to explain the short-lived quality of the results. Presumably, the experimenter's response-contingent comments would not have acquired strong reinforcing properties, since they became, perforce, rather redundant signals. In other words, because the learning task proved to be extremely difficult, the experimenter was required to administer social reinforcements with great frequency (Mean trials to criterion on List II was 25; cf. Atkinson & Robinson, 1961). Consequently, the meaningfulness and psychological effects of the experimenter's repeated evaluations would be expected to decline in relationship to their frequency (Cairns, 1963).

To summarize, the abrupt change from nonevaluation to social reward and, especially, to social punishment served inadvertently to distract and disconcert both the HD and HD-A subjects—but not to the same extent. Moreover, because of the complexity of the task, the experimenter's verbal reinforcements rapidly lost much of the motivational effects they had held on List II. Thus, for all intents and purposes, by List III and thereafter, the four reinforced groups may well have been performing under conditions approaching nonreinforcement.

Despite these unplanned-for effects, the results did suggest that dependency and dependency-anxiety are relevant determinants of patients' responsivity to social evaluations. Indeed, the view that responsiveness to different social reinforcements is contingent upon the individual's dependency needs has now been demonstrated with four widely divergent populations: schizophrenic adults, college students (Cairns & Lewis, 1962), delinquent adolescents (Cairns, 1959), and young children (Gewirtz & Baer, 1958).

Another implication concerns the fact that while introductory or practice lists are typically administered under nonevaluation, they are not always included in the analysis (e.g., Gladis & Wischner, 1962). In view of the learning impairments produced by shifting patients from conditions of nonevaluation to social evaluation, improvement scores based only on the differences in learning between the first and last *reinforced* lists may sometimes be misleading. Unless the practice list performances are also taken into account, the subjects encountering the initially disruptive condition of punishment would eventually appear to have improved more than subjects who encountered rewards, and much more than subjects who continued under nonevaluation.

The results of the ward behavior classifications might also have some bearing on this discussion. Although of uncertain validity, these ratings by the ward administrators suggested that in a neuropsychiatric hospital the proportion of HD-As to HDs was approximately two to one. Such a disproportionate representation of dependency-anxious personalities in previous patient samples could explain in part why schizophrenics have generally been found to perform more effectively under punishment than under reward. In any event, it seems likely that this issue of punishment versus reward will not be resolved satisfactorily until the influence of personality on learning has been considered in its entirety.

REFERENCES

ATKINSON, RITA L., & ROBINSON, NANCY M. Paired-associate learning by schizophrenic and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, **62**, 322–326.

BANDURA, A., & WALTERS, R. H. *Adolescent aggression*. New York: Ronald Press, 1959.

BERNADIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *Journal of Consulting Psychology*, 1957, **21**, 63–67.

CAIRNS, R. B. The influence of dependency-anxiety on the effectiveness of social reinforcers. Unpublished doctoral dissertation, Stanford University, 1959.

CAIRNS, R. B. Antecedents of social reinforcer effectiveness. Paper read at Biennial Conference of SRCD, Indianapolis, April 1963.

CAIRNS, R. B., & LEWIS, M. Dependency and the reinforcement value of a verbal stimulus. *Journal of Consulting Psychology*, 1962, **26**, 1–8.

D'ALESSIO, G. R., & SPENCE, JANET T. Schizophrenic deficit and its relation to social motivation. *Journal of Abnormal and Social Psychology*, 1963, **66**, 390–393.

EGGER, M. D., & MILLER, N. E. When is a reward reinforcing?: An experimental study of the information hypothesis. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 132–137.

FISCHER, E. H. Task performance of chronic schizophrenics as a function of verbal evaluation and social proximity. *Journal of Clinical Psychology*, 1963, **19**, 176–178.

FROMM-REICHMANN, FRIEDA. Psychotherapy of schizophrenia. In D. M. Bullard (Ed.), *Psychoanalysis and psychotherapy: Selected papers of Frieda Fromm-Reichmann*. Chicago: Univer. Chicago Press, 1959.

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology*, 1958, **56**, 49–56.

GLADIS, M., & WISCHNER, G. J. Schizophrenic and normal response patterns to "aversive" and "neutral" associations in two paired-associate paradigms. *Journal of Abnormal and Social Psychology*, 1962, **64**, 249–256.

KARRAS, A. The effects of reinforcement and arousal on the psychomotor performance of chronic schizophrenics. *Journal of Abnormal and Social Psychology*, 1962, **65**, 104–111.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

TAYLOR, JANET. Drive theory and manifest anxiety. *Psychological Bulletin*, 1956, **53**, 303–320.

(Received November 23, 1964)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 419–425

ANTICIPATION INTERVAL AND AGE DIFFERENCES IN VERBAL LEARNING

DAVID ARENBERG[1]

Gerontology Branch, National Heart Institute, Baltimore City Hospitals, Baltimore, Maryland

Age differences were demonstrated, and the age differences were greater at the shorter anticipation interval than at the longer in 2 paired-associate studies. Earlier work had suggested that this age-pace interaction was not a true learning effect but a performance deficit due to the difficulty the older person has in responding rapidly. The second study provided evidence that relatively few of the errors made by the old Ss at the fast pace would be attributed to insufficient time to emit a learned response. Even when Ss were permitted to take as much time as needed to respond, the age-pace interaction was found. Several possible explanations for such a learning deficit were noted.

Several studies have indicated that the verbal-learning performance of old subjects is inferior to that of young adults, but before 1960 the presentation rate (pace) had not been evaluated as an independent variable in gerontology (cf. reviews by Jerome, 1959; Kay, 1959; Welford, 1958). Recently, the pace variable has been explored in a group of studies (Canestrari, 1963; Crovitz, 1960; Eisdorfer, Axelrod, & Wilkie, 1963) at Duke University; and as predicted, differences between old and young groups were found to be greater at faster paces than at slower paces. In Canestrari's (1963) paired-associate study each subject performed at three different pace conditions: (*a*) fast pace which was 1.5 seconds for the anticipation interval and 1.5 seconds for the inspection interval, (*b*) slow pace which was 3.0 seconds for each interval (6.0 seconds per item), and (*c*) self-paced in which the subjects were permitted to extend the time whenever they believed it necessary. The largest age difference occurred at the fast pace and the smallest age difference at the self-paced condition. Eisdorfer et al. (1963) conducted two serial-anticipation studies from which they concluded that the prevalence of errors among old subjects at a fast pace results from insufficient time to produce learned responses. Their first study, using two age groups each divided into three pace groups, indicated that the age-pace interaction for total errors was attributable to the omission errors. Commission errors showed little variation between age groups or among the three pace groups; only the omission errors for the old groups seemed to contribute to the age-pace interaction in which the largest age difference was found at the fastest pace and the smallest age difference at the slowest pace. In their second study, two groups of old subjects were switched from one pace to the other for the eighth and ninth acquisition trials. The original slow-pace group increased their errors during the eighth and ninth trials (which were fast), whereas the original fast-pace group decreased their omission errors on the eighth and ninth trials (which were slow).

Comparisons of omission and commission errors have suggested that the high error scores of the old subjects at fast paces were performance rather than learning deficits; but, as Eisdorfer et al. noted, such an analysis of types of errors cannot provide a definitive answer to the performance-learning question. Some errors in paced trials may be due to insufficient time to respond; others are failures to learn. Errors of omission include both learning and performance errors; therefore, the omission-commission analysis does not separate learning from performance. Direct resolution of the learning-performance problem in verbal learning requires measures of errors which are not attributable to insufficient time to respond. To provide such

---

[1] The assistance of Patricia Lamb, Richard Mathias, and Don McQ. Reynolds in collecting, scoring, and analyzing data is gratefully acknowledged.

TABLE 1

MEANS FOR AGE AND WAIS VOCABULARY RAW SCORE

	Age		Vocabulary	
	Fast pace	Slow pace	Fast pace	Slow pace
Study I				
Old	70.2	70.2	66.6	66.5
Young	35.9	35.8	64.4	63.7
Study II				
Old	66.0	65.8	42.7	44.9
Young	19.8	19.6	42.5	41.1

Note.—In each cell, N = 16.

measures the second of the two studies reported here included self-paced trials (without feedback) alternated with paced trials. In addition, for both paired-associate studies, the anticipation interval (opportunity-to-respond time) was independent of the inspection interval (stimulus displayed together with the response), and only the anticipation interval was varied. In this way, differences resulting between pace groups are not attributable to additional time to view the material to be learned.

It was hypothesized in both studies that: (*a*) the old subjects would make more errors than the young, and (*b*) the age difference would be greater at the faster than at the slower pace. In addition, it was hypothesized in the second study that the pattern of results for self-paced errors would be similar to the pattern of paced errors despite the fact that each subject could take as much time as he needed to respond; that is, a pace difference between the old groups and an age-pace interaction was predicted for self-paced errors as well as for paced errors.

STUDY I

Method

Subjects. The subjects were male volunteers participating in a longitudinal program who were selected for this cross-sectional analysis on the basis of age. Each participant was randomly assigned to the fast or slow pace. When data were available for 16 men in each pace group who were 40 years of age or below (29–40), the sample for this study was closed. The oldest 16 men in each pace group at that time then made up the two old groups (63–77). Virtually all of the 64 subjects held (or had retired from) academic, scientific, technical, or administrative positions primarily in government service. All 32 young subjects had some college education and almost all had a degree. Although one old subject did not have a high school education, over half of the old group had an advanced degree. Scores of all 64 subjects on the Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS) exceeded the mean of the national standardization sample. Means for age and WAIS Vocabulary raw score appear in Table 1.

Material and procedure. The eight original items of the Gladis and Braun retroaction study (1958) were used. The stimulus component of each item consisted of two consonants, and the response component was a familiar, two-syllable adjective; for example, TL INSANE, SN COMPLETE, etc.

A card-changing instrument (Hunter Cardmaster) with independent time controls for each of 3 intervals (left shutter open, both shutters open, both shutters closed) was used to present the material. The anticipation interval, during which the 2 consonants were exposed by opening the left shutter, was 1.9 seconds for the fast pace and 3.7 seconds for the slow pace. For both pace groups, the inspection interval, during which consonants together with the word were exposed by opening both shutters, was 1.9 seconds. Also the interitem interval, during which both shutters were closed, was 1.8 seconds for both pace groups. Five different orders of the 8 items were repeated until 1 errorless trial or 52 trials occurred. Repetitions of any item were always separated by at least 2 other items, and no item immediately followed the same item more than once throughout the 5 orders.

The subjects performed individually and called out their responses which were recorded by the experimenter. The dependent measure was the total number of errors. Guessing was encouraged.

Results

The mean number of errors for each age-pace combination appears in Table 2. Mean age differences were 123.0 errors at the fast pace and only 27.3 errors at the slow pace. This interaction effect and the main effects of age and pace were evaluated by analysis of variance. To test the unidirectional hypotheses, F values were converted to t; all

TABLE 2

MEANS AND STANDARD DEVIATIONS OF ERROR SCORES (STUDY I)

	M		SD	
	Fast pace	Slow pace	Fast pace	Slow pace
Old	175.1	84.2	105.6	62.4
Young	52.1	56.9	34.2	56.6

Note.—In each cell, N = 16.

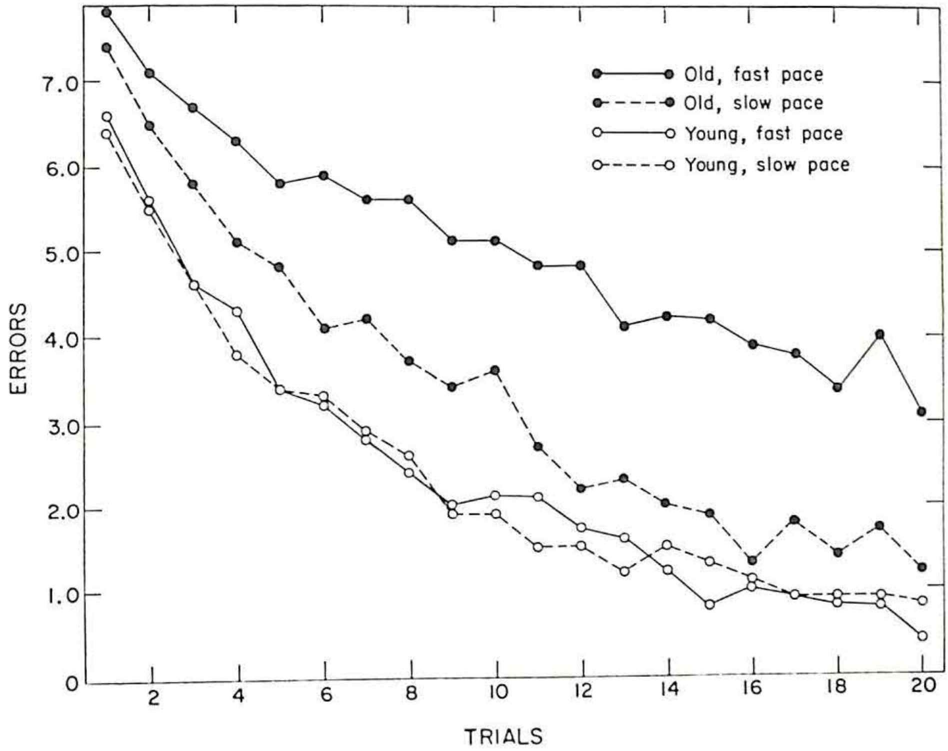


FIG. 1. Mean errors for all four age-pace groups in Study I plotted over the first 20 trials.

effects were significant beyond the 0.01 level. Due to the substantial heterogeneity of variance (see standard deviations in Table 2) and the relation between means and variances, scores were transformed to logarithms; but the resulting analysis of variance was essentially the same (all *t* values significant beyond 0.03). The within-groups correlation (Lindquist, 1953, p. 321) of −0.20 between WAIS Vocabulary scores and error scores was not significantly different from zero at the 0.05 level.

Inspection of Figure 1 indicates that the group differences emerged quite early in the acquisition trials.

Were the higher error scores of the old group at the fast pace due to difficulty in learning at that pace, or was the material learned but not manifested because the subjects could not respond in the short anticipation interval? The second study was designed to determine whether similar results would be obtained for a less educated sample with lower WAIS Vocabulary scores; and, if so, whether the relatively poorer performance of the older subjects at the faster pace is primarily a learning or a response deficit.

STUDY II

Method

Subjects. Unemployed men actively seeking work at one of two employment agencies of the State of Maryland were paid to participate. In order to maximize the probability of recruiting men who could read without difficulty, at least a sixth-grade education was required; but this initial screening was not entirely successful. Therefore, the subjects with a raw score on the WAIS Vocabulary below 20 were excluded. The age ranges were 18–21 for the young groups and 60–77 for the old. There were 16 subjects in each of the four age-pace groups, and the 64 subjects were randomly assigned to the fast or slow pace. Means for age and WAIS Vocabulary raw score appear at the bottom of Table 1.

Material and procedure. Preliminary data had indicated that the eight-item list used in Study I was too difficult for the subjects in Study II. Six items were constructed using pairs of consonants from Study I as stimuli; responses were familiar five-letter words with the second and fourth letters vowels: TL COVER, BV MUSIC, FY LEVEL, DM TIGER, KZ WAGON, WG SUGAR.

To provide measures of errors which are not attributable to insufficient time to respond, self-paced test trials were alternated with paced acquisition trials throughout the session for every subject. This procedure yielded two performance measures at each stage of learning: (*a*) self-paced—the number of errors for six items during which the subject was permitted to take as much time as he needed to re-

TABLE 3

DISPLAYS AND DURATIONS FOR A SELF-PACED AND A PACED TRIAL (STUDY II)

	Display	Duration (in seconds)
Self-paced	TL	2.0
	(no display)	[a]
	BV	2.0
	(no display)	[a]
	FY	2.0
	(no display)	[a]
	KZ	2.0
	(no display)	[a]
	WG	2.0
	(no display)	[a]
	DM	2.0
	(no display)	[a]
Paced	KZ	1.9 or 3.7[b]
	KZ WAGON	1.9
	(no display)	1.8[c]
	FY	1.9 or 3.7[b]
	FY LEVEL	1.9
	(no display)	1.8[c]
	BV	1.9 or 3.7[b]
	BV MUSIC	1.9
	(no display)	1.8[c]
	DM	1.9 or 3.7[b]
	DM TIGER	1.9
	(no display)	1.8[c]
	TL	1.9 or 3.7[b]
	TL COVER	1.9
	(no display)	1.8[c]
	WG	1.9 or 3.7[b]
	WG SUGAR	1.9
	(no display)	1.8[c]

[a] These intervals a minimum of 1.8 seconds, but not terminated until the subject responded with word or said, "No guess."

[b] These intervals all 1.9 seconds for the subjects assigned to fast pace and all 3.7 seconds for the subjects assigned to slow pace.

[c] Apparatus required a minimum 1.8 seconds to change items; minimum change-time also accounted for 1.8 seconds minimum for no-display intervals in self-paced trials.

spond, and (*b*) paced—the number of errors for six items when the anticipation interval was determined.

Timing of the paced trials was the same as in Study I. The anticipation interval during the paced trials was 1.9 seconds for the subjects assigned to the fast pace group and 3.7 seconds for subjects assigned to the slow pace group. In the self-paced trials for all subjects only the pairs of consonants were displayed. A pair of consonants was exposed by opening both shutters for 2.0 seconds followed by at least 1.8 seconds with both shutters closed. The next pair of consonants was not presented until the subject responded or said, "No guess." At no time during the self-paced test trials were the response words shown. The material displayed and the durations of each display are listed in Table 3 for a self-paced test trial followed by a paced acquisition trial.

Prior to each paced trial, the subject was instructed to respond before the word appeared; prior to each self-paced trial, the subject was instructed to take as much time as he really needed to respond. Five item orders were used as follows: Order 1 for the first self-paced trial, Order 2 for the first paced trial, Order 3 for the second self-paced trial, Order 4 for the second paced trial, Order 5 for the third self-paced trial, Order 1 for the third paced trial, etc., until Order 5 was used for the fifth paced trial. Then the sequence was repeated until four consecutive errorless trials (two self-paced and two paced) occurred or for a total of 39 self-paced and 39 paced trials.

TABLE 4

MEANS AND STANDARD DEVIATIONS OF ERROR SCORES (STUDY II)

	Paced performance		Self-paced performance	
	Fast-pace group	Slow-pace group	Fast-pace group	Slow-pace group
M				
Old	121.6	59.4	107.9	56.7
Young	21.4	16.4	16.4	15.8
SD				
Old	62.4	39.1	57.9	37.8
Young	22.4	16.9	17.9	14.6

Note.—In each cell, $N = 16$.

Just as in Study I, subjects performed individually and called out their responses which were recorded by the experimenter; the subjects were encouraged to guess.

Results

Means and standard deviations for paced and self-paced errors appear in Table 4. An

TABLE 5

ANALYSES OF VARIANCE FOR ERRORS IN STUDY II

Source (all errors)	*df*	*MS*	*t*
Age (A)	1	151,800.5	7.33
Pace (B)	1	28,322.0	3.17
A × B	1	23,220.1	2.87
Error (b)	60	2,824.1	
Paced, self-paced (P, S-P)	1	968.0	4.10
P, S-P × A	1	231.1	2.00
P, S-P × B	1	465.1	2.84
P, S-P × A × B	1	91.1	1.26
Error (w)	60	57.6	
Source (self-paced errors)	***df***	***MS***	***t***
A	1	70,092.6	7.27
B	1	10,764.1	2.85
A × B	1	10,201.0	2.77
Error	60	1,326.0	

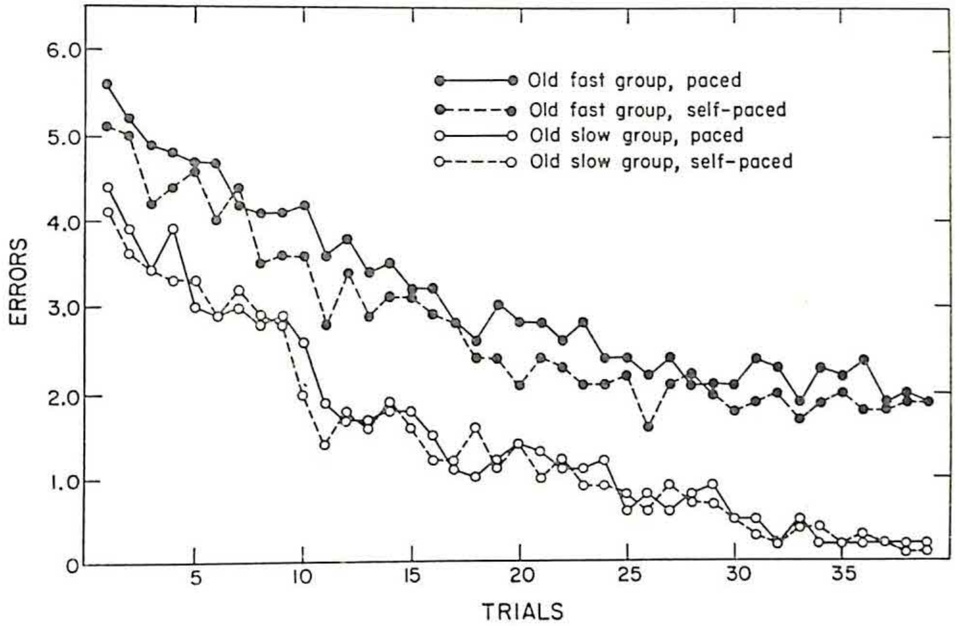


FIG. 2. Means of paced and self-paced errors for both old groups in Study II plotted over all trials.

analysis of variance (Type III, Lindquist, 1953) indicated the main effects of age and pace and the age-pace interaction were all statistically significant (see Table 5) similar to the results of Study I. In addition, the within-subject aspect of the analysis showed statistical significance for the main effect of pacing versus self-pacing, and the interactions of this variable with age and with pace; but the triple interaction was not significant. The age differences were greater for the fast-pace groups than for the slow, but this pattern appeared to be similar for paced and self-paced errors.

An analysis of only the self-paced errors more clearly answered the primary question. If the older subjects at the fast pace cannot respond in time, but learn as well as the old subjects at the slow pace, then analysis of the self-paced errors should show no age-pace interaction effect and no difference between old groups at the two paces. If, however, the old subjects do not learn as well at the fast pace, the age-pace interaction should be maintained in the analysis of self-paced errors and the old group at the fast pace should make more self-paced errors than the old group at the slow pace. The analysis indicated that the age-pace interaction was maintained and the fast-pace old subjects made significantly more errors ($p < 0.01$) than the slow-pace old group even when permitted to take as much time as needed to respond. The plot of paced and self-paced errors for both old groups (see Figure 2) more clearly illustrated that throughout all trials, the self-paced performance of the fast-pace group was more like its own paced performance than

TABLE 6

ANALYSES OF VARIANCE FOR LOG ERRORS IN STUDY II

Source (all errors)	*df*	*MS*	*t*
Age (A)	1	18.0676	6.47
Pace (B)	1	1.1156	1.61
A × B	1	1.3183	1.75
Error (b)	60	0.4310	
Paced, self-paced (P, S-P)	1	0.0549	2.05
P, S-P × A	1	0.0001	<1
P, S-P × B	1	0.0218	1.29
P, S-P × A × B	1	0.0108	<1
Error (w)	60	0.0131	
Source (self-paced errors)	*df*	*MS*	*t*
A	1	8.9700	6.46
B	1	0.4128	1.39
A × B	1	0.7832	1.91
Error	60	0.2147	

like the self-paced scores of the slow-pace group.

Due to the heterogeneity of variances and the relation between means and variances (see Table 4), scores were transformed to logarithms. The results (see Table 6) were essentially the same for the effects of age, paced versus self-paced, and the age-pace interaction, but the main effect of pace and the double interactions within subjects were no longer statistically significant. In the analysis of transformed self-paced errors, the age effect and the age-pace interaction effect were maintained and the difference between the two old groups was again statistically significant ($p < 0.01$).

DISCUSSION

The results of Study I were consistent with the findings and the performance-decrement interpretation of Eisdorfer et al. (1963). Evidence was found in Study II that for old subjects learning is impaired when the anticipation interval is short. The fact that differences between paced and self-paced errors were small suggested that few of the paced errors, even for the old group at the fast pace, can be attributed to insufficient time to emit a correctly learned response.

Welford [2] has suggested that the process of search may account for age interactions involving underlearned responses as well as interactions involving overlearned responses. Suci, Davidoff, and Surwillo (1960) reported that age differences in time to make overlearned responses increased with increases in number of choices. These findings may be thought of as relatively greater increases in search time required by old subjects as the pool of material to be searched increases. A similar search deficit with age may also occur when information storage is incomplete, that is, when responses are underlearned. Presumably search takes longer for older subjects resulting in poorer performance than for young subjects (particularly at a short anticipation interval), and cumulative search failures may also be detrimental to learning as well as to performance.

A related hypothesis is that active responding is conducive to learning, and the short anticipation interval during acquisition trials decreases the probability of a correct response for old subjects. If this is so, performance decrements would occur early and learning decrements would follow for old subjects due to the cumulative failure to respond at the short anticipation interval. Kay and several others have observed that older subjects frequently commit an error early in acquisition and then repeat that error on many subsequent trials despite disconfirmation of their incorrect response (cf. Kay, 1959). It seems that such subjects have learned a response but a wrong one. If what response the old person makes determines to some extent what he learns, then perhaps whether a response is made determines to some extent whether learning occurs.

It is also possible that subjects in the slow-paced groups rehearse more and that additional time to rehearse is more beneficial to old subjects than to young. This possibility cannot be ruled out, but there was no evidence available that the old subjects in Study II were rehearsing during the anticipation intervals of the paced trials.

Another possibility is that interitem interference is greater at a fast pace because the next item appears sooner and there is less time for the undisrupted trace of an item to enter long-term storage (cf. discussion of retention in Welford, 1963). If the trace in an older person is more susceptible to interference from intrusion of the next item, a fast pace should be more detrimental to learning for him.

Hypotheses involving search, active responding, interference, and rehearsal can be investigated by various procedural techniques including item massing (Greeno, 1964), prompting (Cook & Spitzer, 1960), and part-learning, as well as by comparing age-pace interactions in factorial studies with more than one pace variable (anticipation, inspection, and interitem intervals).

REFERENCES

CANESTRARI, R. E., JR. Paced and self-paced learning in young and elderly adults. *Journal of Gerontology*, 1963, 18, 165–168.

COOK, J. O., & SPITZER, M. E. Supplementary report: Prompting versus confirmation in paired-

---

[2] Welford, personal communication.

associate learning. *Journal of Experimental Psychology*, 1960, 59, 275–276.

CROVITZ, ELAINE K. Rote learning in the aged. Unpublished master's thesis, Duke University, 1960.

EISDORFER, C., AXELROD, S., & WILKIE, FRANCES L. Stimulus exposure time as a factor in serial learning in an aged sample. *Journal of Abnormal and Social Psychology*, 1963, 67, 594–600.

GLADIS, M., & BRAUN, H. W. Age differences in transfer and retroaction as a function of intertask response similarity. *Journal of Experimental Psychology*, 1958, 55, 25–30.

GREENO, J. G. Paired-associate learning with massed and distributed repetition of items. *Journal of Experimental Psychology*, 1964, 67, 286–295.

JEROME, E. A. Age and learning—experimental studies. In J. E. Birren (Ed.), *Handbook of aging and the individual*. Chicago: Univer. Chicago Press, 1959.

KAY, H. Theories of learning and aging. In J. E. Birren (Ed.), *Handbook of aging and the individual*. Chicago: Univer. Chicago Press, 1959.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton-Mifflin, 1953.

SUCI, G. J., DAVIDOFF, M. D., & SURWILLO, W. W. Reaction time as a function of stimulus information and age. *Journal of Experimental Psychology*, 1960, 60, 242–244.

WELFORD, A. T. *Ageing and human skill*. London: Oxford Univer. Press, 1958.

WELFORD, A. T. Social, psychological, and physiological gerontology—an experimental psychologist's approach. In R. H. Williams, C. Tibbitts, & Wilma Donahue (Eds.), *Processes of aging*. New York: Atherton Press, 1963.

(Received January 7, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 426–431

MEANING DIMENSIONS AND MALE-FEMALE VOICE PERCEPTION IN SCHIZOPHRENICS WITH GOOD AND POOR PREMORBID ADJUSTMENT[1]

M. J. DONOVAN[2] AND WARREN W. WEBB

Vanderbilt University

This study explored certain aspects of perceptual distortion in schizophrenia. Among schizophrenics with a good premorbid adjustment, schizophrenics with a poor premorbid adjustment, and normals, auditory perception of spoken material was compared as a function of (a) sex of voice and (b) dimensions of meaning. The Ss listened to successive presentations of a recorded spoken word, mixed with decreasing amounts of masking noise, until their recognition thresholds were established. Schizophrenics with poor premorbid adjustment responding to the female voice had significantly higher thresholds than those responding to the male voice. This finding did not occur in schizophrenics with good premorbid adjustment.

Studies separating male schizophrenics on the basis of good and poor premorbid adjustment have contributed to the clarification of the schizophrenic disorder by emphasizing the different roles of the father and mother (Garmezy & Rodnick, 1959). The father tends to be the more dominant parent in the families of the good premorbid adjustment group "goods," while the mother tends to be more dominant in the families of the poor premorbid adjustment group "poors." Direct studies of the parents of such patients (Farina, 1960) support this conclusion.

Early family experiences which feature highly deviant parental behavior may influence sensitivity to the parental figures and help shape the pattern of meaning dimensions which an individual acquires. Engelhart (1959) had normals, goods, and poors rate such concepts as "love" and "scold" on the three bipolar meaning dimensions of the semantic differential (Osgood, Suci, & Tannenbaum, 1957). He found that poors rated negatively toned (censuring) concepts as more potent and active than did goods and normals. This suggested to Garmezy and Rodnick (1959) that poors "are prepared to assign to such cues . . . meanings congruent with the patient's omnipresent perception of threat in his external environment . . . [p. 461]." But Engelhart also observed that poors "uniformly tended to shift their judgments in a more positively evaluated direction . . . [p. 2655]" when the words "mother" or "father" were attached to the initial concepts. Because his data were obtained in the form of ratings, it is possible that the poors felt vulnerable to the potential threat of openly expressing more negative attitudes regarding the role characteristics of the parents and altered their judgments defensively.

Webb (1964) attempted to circumvent such presumed conscious distortion of meaning dimensions by comparing normals, goods, and poors for differences in "selective attention" to words which were fairly extreme (as previously rated by judges) on each of Osgood's three bipolar dimensions. The words were presented visually, in continuously changing clusters under a diffusing glass. Under these conditions, normals and goods were similarly alert to "good" words and relatively insensitive to "inactive" words, while poors appeared not to be differentially sensitive to any of the meaning categories. This was interpreted as indicating a minimal degree of intrinsic motivational interests of poors under the impersonal conditions of the experiment.

---

[1] This paper is a condensation of the senior author's doctoral dissertation, completed at Vanderbilt University in 1963 under the junior author's supervision. The authors wish to acknowledge their indebtedness to M. Ray Mickey, Jr., for his extensive statistical assistance. Appreciation is also expressed to Ronald Wendahl and Eliot Rodnick for the use of facilities in producing stimulus materials, and to Rue Cromwell for assistance with the manuscript.

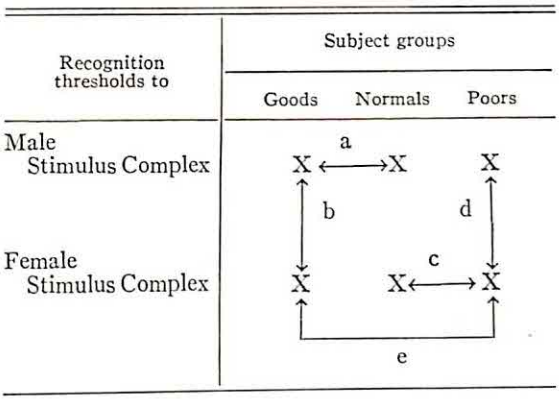
[2] Now at Metropolitan State Hospital, Norwalk, California.

This study was designed to investigate the relationships between premorbid adjustment, mother versus father as a stimulus complex, and dimensions of meaning. This was accomplished by having the subject listen to repeated, increasingly clear, recorded stimulus words. The subject who heard the words spoken by a male voice was confronted with the portrait of a father figure. Similarly, the subject who heard the words spoken by a female voice was confronted with the portrait of a mother figure. No threat or censure or taboo words were incorporated in the stimulus conditions.

While it was expected, in general, that the more psychopathological relationships would be reflected in higher thresholds, formal directional hypotheses on the basis of the literature were regarded as unwarranted. Several hypotheses were stated and are presented schematically in Table 1. For example, it was expected that the recognition thresholds of the poors responding to the female stimulus complex and the recognition thresholds of the goods responding to the male stimulus complex would differ from the thresholds of normals responding to either (c and a, respectively). The literature did not support hypotheses for differences beyond those shown in Table 1, nor did it warrant specific hypotheses in regard to the meaning categories. The experimental design did provide for an analysis of such effects.

TABLE 1

HYPOTHESIZED DIFFERENCES IN RECOGNITION THRESHOLDS BETWEEN SUBJECT GROUPS



METHOD

Subjects

The three groups (poors, goods, normals) were white males, between the ages of 18 and 44 years, with at least an eighth-grade education. The normals were psychiatric technicians assigned to the male wards of Metropolitan State Hospital, Norwalk, California. The poors and goods, diagnosed as schizophrenic, were patients at the same hospital. Only relatively acute cases were used; length of hospitalization averaged 3.9 months. The patients did not receive tranquilizing medication for a period of 3 weeks prior to the experiment.

Hospital records were screened and patients interviewed to obtain subjects who met the following criteria: (*a*) a sufficient social history was available to permit a valid rating on the Phillips Scale of Premorbid Adjustment (1953), (*b*) no evidence of extra-schizophrenic pathology (drug addiction, alcoholism, mental retardation), (*c*) the subjects were willing to cooperate in the test procedures and were able to comprehend instructions, and (*d*) the subjects were reared in a home having both a male and a female authority figure. After

TABLE 2

RANGES AND MEANS OF SIX SUBJECT GROUPS ON FOUR CONTROL VARIABLES: AGE, EDUCATION, VOCABULARY, AND PHILLIPS (PREMORBID ADJUSTMENT) SCORES

Subject group	Sex of voice	Age		Education		Vocabulary score[a]		Phillips score	
		M	Range	*M*	Range	*M*	Range	*M*	Range
Poors	male	30.0	18–40	11.5	8–16	28.4	20–36	21.7	20–26
	female	24.2	18–40	11.7	9–16	28.1	23–37	21.6	20–26
Goods	male	32.8	22–44	12.8	9–16	28.9	22–38	12.5	8–15
	female	31.4	18–40	12.8	9–16	29.7	23–37	11.7	7–15
Normals	male	29.1	21–37	13.0	10–16	30.6	25–34	9.0	4–15
	female	27.1	19–38	14.0	12–18	31.6	27–36	8.5	3–11
Total		29.1	18–44	12.6	8–18	29.6	20–38		

[a] Raw scores, Shipley-Hartford vocabulary test.

this, the patients were scored on the Phillips Scale of Premorbid Adjustment, and groups were constituted of 20 poors, 20 goods, and 20 normals. As may be seen from Table 2, they did not differ significantly in age, education, or vocabulary intelligence (Shipley, 1940).

The subjects in each of the groups were assigned randomly to one of two conditions, identical except for the sex of the stimulus voice and the pictured parental-type figure. Half the poors, goods, and normals were to hear the stimulus words spoken by a male voice; the other half were to hear the same words in the same randomized order spoken by a female voice. Finally, all subjects were screened for auditory acuity in the speech frequency range. There were no systematic differences among groups.

Meaning Dimensions

The Semantic Differential determined the selection of meaningful stimulus words (Osgood, Suci, & Tannenbaum, 1957). The three principal meaning dimensions (value, potency, activity) are represented in this study by six categories of words taken from both ends of the three bipolar scales; *good-bad, strong-weak, active-inactive.* Only those words were used which had been rated in previous studies as fairly extreme on one and only one dimension. In addition, a seventh category of 10 words rated as *neutral* was added to enable more valid interpretations to be made of any differences found between subject-groups on the meaningful categories. In all, 70 words were chosen from two sources: (*a*) an atlas of semantic profiles compiled in a normative study by Jenkins, Russell, and Suci (1958) and (*b*) a list of rated words used by Webb (1964).

The seven lists of words were roughly equated for number of syllables and grade level. All words were below the eighth-grade level (Buckingham & Dolch, 1936). No taboo words were used. All words, as pronounced, had only one generally accepted meaning. Finally, words were chosen which were believed important in interpersonal relationships, and more apt to be used descriptively by one person speaking to another. Most of the words, then, were adjectives.

Preparation of Tapes

The stimulus words were randomized, and complete presentations of a single sequence of the words were recorded in clear voice (without background noise) on two magnetic tapes. One tape was recorded using a male voice; another tape was recorded using a female voice. This was done by phoneticians, who also equated the words for vowel strength.

From each of the clear-voice tapes a master tape was prepared. A tape loop of each word and a random-noise generator was used to produce 9 consecutive, increasingly clear presentations of each stimulus word. Since some words require more masking noise than others (Moser, Michel, & Fotheringham, 1961), an approximate means of equating steps of intelligibility was devised. Using college students as listeners, 2 levels of masking noise were empirically determined. Other levels were then arbitrarily established. One was a point with somewhat higher noise than 2; 2 was the point at which 10% of the students recognized the word; 3 was a point midway in noise level between 2 and 4; 4 was the point at which 50% of the students recognized the word; Points 5, 6, 7, and 8 were equally distributed in noise level between 4 and 9; 9 was complete clarity (no noise).

When completed, the final stimulus tapes consisted of: (*a*) an auditory acuity check, (*b*) instructions to the subjects, (*c*) 3 practice words, and (*d*) the list of 70 randomized stimulus words. On the male-voice tape, all instructions, as well as the stimulus words, were given in a male voice. In the same way, on the female-voice tape, identical instructions were spoken in as closely similar a manner as possible. The speaker in both cases attempted to assume a "neutral" attitude.

A subject's score was established as the number of the presentation at which he made the first of two consecutively correct responses. High scores thus indicate high recognition thresholds.

Procedure

Without prior contact with the experimenter, each subject was brought into a room containing only the research materials. In the manner of a routine testing situation, he was asked to sit at a desk, on which were mounted a loudspeaker, microphone, and a metal box containing a light and a switch. Also on the desk directly in front of the subject was a framed 8 × 10 inch photograph of a middle-aged mother-figure (or father-figure) to match the voice the subject would be hearing. These photographs had been rated "most neutral" in a previous study (Alvarez, 1957).

The subject was told he would receive a light signal and then a stimulus word. When this occurred, he was asked to guess what the word was and then to turn off the light switch. (The light switch allowed a motor as well as a verbal response, in order to offset dwindling motivation and withdrawal.) The light was activated manually by the experimenter 2 or 3 seconds before the presentation of a word and was wired so that the experimenter could switch it off from his position in the adjoining room if the subject failed to do so. This was seldom necessary. Each stimulus word was presented consecutively with increasing clarity up to 9 times and with 7-second intervals. When the subject twice reported it correctly, the presentation of the next word in the random sequence was begun in the same manner. The presentation where he first reported the word correctly was taken as his threshold score. All 70 words were presented in a single testing session of approximately 70 minutes. At the midpoint of the word list, the subject was allowed a 10-minute "coke" break, in which he could walk around, smoke, have a coke, but have minimal interaction with the experimenter.

RESULTS

Analysis of Variance of Scores

The threshold recognition levels of the subjects were analyzed in a 3 × 2 ×7 (mixed) analysis of variance design (Lindquist, 1953), with adjustment level and male versus female stimulus complex as between effects and with word categories as within treatment effects. As may be seen in Table 3, the major prediction of the study was confirmed through the significant interaction of adjustment level and male versus female stimulus complex (A × B). As expected, but of no consequence, adjustment level as a main effect was significant. Also expected, but of no consequence, male versus female stimulus complex as a main effect was not significant. An analysis of covariance, based upon adjustment for age and vocabulary scores (Cochran & Cox, 1957, p. 86), was also computed. Results were similar to the original analysis.

To obtain a more revealing picture of the differences in mean performance between the six subject-groups, adjusted mean scores were used to make multiple comparisons among the means of all subject-groups. The method used was Duncan's new multiple-range test (Edwards, 1960, p. 136). The results are shown in Table 4.

TABLE 3

ANALYSIS OF VARIANCE OF RECOGNITION THRESHOLDS (IN TERMS OF INCREASINGLY CLEAR PRESENTATIONS) AS A FUNCTION OF PREMORBID ADJUSTMENT, MALE VERSUS FEMALE STIMULUS COMPLEX, WORD CATEGORIES, AND WORDS WITHIN CATEGORIES

Source of variation	*SS*	*df*	*MS*	*F*
Between subjects				
Ss adjustment level (A)	1,584.96	2	792.48	42.34*
Male versus female stimulus complex (B)	22.00	1	22.00	1.18
A × B	260.40	2	130.20	6.96*
Error (b)	1,010.79	54	18.72	
Within subjects				
Word categories (C)	268.00	6	44.67	10.59*
Words within categories (D)	2,735.95	63	43.43	10.29*
A × C	161.29	12	13.44	3.18*
A × D	588.02	126	4.67	1.11
B × C	439.49	6	73.25	17.36*
B × D	2,898.60	63	46.01	10.90*
A × B × C	54.52	12	4.54	1.08
A × B × D	562.48	126	4.46	1.06
Error (w)		3,726	4.22	
Total		4,199		

* $p < .005$.

TABLE 4

ADJUSTED MEAN THRESHOLD SCORES
(Corrected for age and vocabulary)

Normals to male stimulus complex	4.190	
Normals to female stimulus complex	4.245	$p < .01$
Goods to female stimulus complex	4.963	
Poors to male stimulus complex	5.143	
Goods to male stimulus complex	5.285	$p < .01$
Poors to female stimulus complex	6.018	

It can be seen by comparison with Table 1 that four of the five (a, c, d, e) expectations were confirmed. The difference between the goods responding to the male stimulus complex and the goods responding to the female stimulus complex approached, but did not reach, an acceptable level of significance.

Response Patterns to Word Categories

As for the meaning dimensions, those significant effects which did occur were vitiated by "within categories" variance.

The adjustment levels by word categories interaction was significant, however, and this was consistent over words within a category (no adjustment levels by "words within categories" interaction). A further examination of the relation of subject-groups to word categories was then made, using group means to rank the categories in three ordered series (Table 5). A pattern of similarity was found in the responses of poors and goods which proved to be significant (Spearman rank correlation coefficient =.82; $p < .05$). There was no significant correlation between poors and normals ($\rho = .07$), or between goods and normals ($\rho = .07$).

Response patterns of poors to the male stimulus complex and poors to the female stimulus complex were then compared with

TABLE 5

RANK ORDER OF WORD CATEGORIES
(according to subject group means)

Rank	Poors	Goods	Normals
1	Good	Good	Active
2	Neutral	Active	Weak
3	Active	Neutral	Inactive
4	Inactive	Bad	Good
5	Weak	Weak	Strong
6	Bad	Inactive	Bad
7	Strong	Strong	Neutral

goods. There was a substantial correlation between the response patterns of the poors to the male stimulus complex and the goods ($\rho = .86$, $p < .05$), but little correlation between the poors responding to the female stimulus complex and the goods ($\rho = .50$, $p > .05$).

DISCUSSION

The Normals did not differ significantly in responding to the male stimulus complex and to the female stimulus complex and thus they provide a dependable base line from which to evaluate the two patient groups. The latter differed significantly from the normals; their performance presumably reflects a schizophrenic deficit which they (poors and goods) share. The poors showed a significant additional deficit as compared with the goods. The question arises, does this deficit reflect merely a greater degree of general schizophrenic impairment (i.e., more of the same thing), or does it reflect qualitative differences in behavioral tendencies as well? There is evidence for each of these alternatives.

When the groups were separated according to the male versus female stimulus complex, the poors responding to the female stimulus complex had the highest thresholds. In responding to the male stimulus complex, however, the poors and goods did not differ significantly. This is consistent with the point of view that the differences in behavioral responses of goods and poors are in part qualitative, in the sense that different kinds of stimulus situations—depending upon whether they have male or female properties—differentiate poors from goods.

The results support Garmezy and Rodnick (1959) and other studies which have shown similar differential behavior toward male-female associated stimuli. The potency of such male-female aspects of interaction is shown here more clearly, perhaps, in that this experiment used no censuring mother or father figures as stimulus material. Threat or censure was not experimentally induced. Nor were taboo words included in the list. In the test situation, there was merely the sound of a female voice (or male voice) speaking words apt to be heard in everyday conversation, coupled with a photograph of a mother figure (or father figure) whose expression had been rated "most neutral." Yet here the poors show higher recognition thresholds even though the situation has only rather nonspecific associations with female figures.

An opposite trend was found in the goods, whose thresholds were higher when exposed to a similar situation involving male figures. In their case, the difference in thresholds between the male stimulus complex and the female stimulus complex was not statistically significant. The direction of the difference, however, considered together with the difference between the two groups of poors, supports the concept of poors being especially disturbed by mother-relationships and of goods, apparently to a lesser extent, by father-relationships.

The effect of the third major variable—meaning dimensions—is not a dominant one, yet it poses some interesting questions. Poors and goods responded to the different categories of meaning in a similar pattern, as contrasted to the normals (Table 5). Again, the similarity of the two patient groups possibly reflects certain behavioral tendencies associated with schizophrenic disorders in general. However, when the response patterns of poors responding to the male stimulus complex and poors responding to the female stimulus complex were compared with goods the former correlated significantly with the goods, while the latter did not. Thus when poors manifested no greater deficit in performance than goods, their response patterns were similar; when poors showed a significant rise in thresholds, the pattern of their responses changed. Again, the performance of the poors responding to the female voice appears to be qualitatively different.

Schizophrenics had lower recognition thresholds for concepts associated with positive or neutral affect (*good, active, neutral* words) and higher thresholds for concepts having negative connotations (*bad, weak, inactive* words). The *strong* dimension—considered to have positive cultural connotations—is the singular exception to the above pattern. The highest thresholds of both poors and goods were found in this category. This may have been a random effect; a meaningful explanation, however, would be that the schizophrenic's apparent aversion to strong words may reflect a lower self-concept—the result of lifelong frustrations and rejection experiences. Thus the schizophrenics may have come to avoid—as alien to the self—the perception of concepts associated with potency. Another meaningful accounting of the high threshold to strong words would be that the schizophrenic has associated potency concepts with threatening parental authority and related punitive experiences.

Only one pattern was discernible in the responses of the normals. They appeared to facilitate perception of all meaningful words in preference to the more abstract character of the neutral words. Poors and goods, on the other hand, appeared to facilitate perception of neutral words. This was probably but a reflection of greater avoidance tendencies toward the negatively toned concepts.

The response patterns found in this study need to be substantiated by further research. If similar patterns appear consistently, a more extensive use of meaning dimensions in studying schizophrenic behavior would be encouraged. The major impression from the present study, however, is that sex of the stimulus complex, in particular combination with goods and poors, exerted a predominant effect compared to meaning dimensions. The differences between adjustment-level and stimulus-complex combinations were consistent across all categories, regardless of meaning. The high thresholds found among poors responding to the female stimulus complex tend to support the body of data which contends that poors, identified with premorbid sexual and social maladjustment, have been more exposed to deviant maternal behavior than have goods.

REFERENCES

ALVAREZ, R. R. A comparison of the preferences of schizophrenic and normal subjects for rewarded and punished stimuli. Unpublished doctoral dissertation, Duke University, 1957.

BUCKINGHAM, R. R., & DOLCH, E. W. *A combined word list.* Boston: Ginn, 1936.

COCHRAN, W. G., & COX, G. M. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

ENGELHART, R. S. Semantic correlates of interpersonal concepts and parental attributes in schizophrenia. *Dissertation Abstracts,* 1959, **19,** 2655–2656.

FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1960, **61,** 31–38.

GARMEZY, N., & RODNICK, E. H. Premorbid adjustment performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *Journal of Nervous and Mental Disease,* 1959, **129,** 450–466.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *American Journal of Psychology,* 1958, **71,** 668–694.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton-Mifflin, 1953.

MOSER, H. M., MICHEL, J. F., & FOTHERINGHAM, W. C. Intelligibility and confusability of vowels and diphthongs under various conditions of quiet and noise. *USAF CCDD tech Note,* 1961, No. 61-33.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *Journal of Nervous and Mental Disease,* 1953, **117,** 515–525.

SHIPLEY, W. S. *Shipley-Institute of Living scale for measuring intellectual impairment.* Hartford, Conn.: Institute of Living, 1940.

WEBB, W. W. Premorbid adjustment and the perception of meaning. *Perceptual & Motor Skills,* 1964, **17,** 762.

(Received January 11, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 432–441

PAROLE OUTCOME AS PREDICTED FROM THE CPI, THE MMPI, AND A BASE EXPECTANCY TABLE[1]

HARRISON G. GOUGH
University of California, Berkeley

ERNEST A. WENK
California Department of Corrections, Deuel Vocational Institution

AND VITALI V. ROZYNKO
California Department of Mental Hygiene, Mendocino State Hospital

Parole outcomes for 183 violators and 261 nonviolators (initial sample), and 130 violators and 165 nonviolators (cross-validating sample) were forecast from demographic and personality inventory data. The best single source of prediction in the cross-validating sample was a demographic Base Expectancy index. The best combination was given by an equation including this index, the socialization, self-control, and communality scales of the CPI with positive weightings, and the social presence scale of the CPI with a negative weighting. Conceptual analysis of the equation from the CPI, MMPI, and CPI + MMPI revealed a clear dimension of psychological meaning, one having significance for normal as well as for deviant behavior.

The prediction of success or failure on parole is a problem which has long been of interest to behavioral scientists. Beginning with the pioneering efforts of Hornell Hart (1923), who first introduced the "experience table" as a technique of prediction, there have been many attempts to resolve the substantive and methodological issues which are involved in this problem. General historical surveys (Monachesi, 1950; Schuessler, 1954) and technical analyses (Glaser, 1962; Gough, 1962) are available, but a few comments on important studies and landmarks should nevertheless be made.

Following Hart's specification of the methodological principles underlying the experience table, two highly significant explorations of the method were undertaken by Burgess (1928) and the Gluecks (1930). A lesser known but nonetheless important application was published by Vold (1931). In the studies by Burgess and Vold systematic surveys were made of demographic and life history items to discover those which would differentiate between violators and nonviolators on parole. In both studies items were found which did so differentiate, and which when assembled into an "experience table" gave rise to a total score which was quite effective in distinguishing between men who were successful or unsuccessful on parole. The study by the Gluecks applied the same technique to the differentiation of delinquents from nondelinquents.

During the ensuing years the experience table method has been cross-checked by many researchers, sometimes with unfavorable results (e.g., Barron, 1962; Gillin, 1950) but more often with positive outcomes (cf. Glaser, 1962; Grant, 1962; Schuessler, 1954). The Gluecks also continued their own work, publishing many studies which have become classics in the fields of prediction and criminology (see especially *Unraveling juvenile delinquency*, 1950).

In the early and middle 1950s two developments of note occurred in the area of experience table methodology. First, an important new table was introduced (Mannheim & Wilkins, 1955), and second, the concepts of expectancy and inverse probability began to be applied to the evaluations of the utility of prediction (Glaser, 1955; Meehl, 1954). From these analyses it became apparent that although the best of the tables, such as those of the Gluecks (1950, 1956) Ohlin (1951) and Mannheim and Wilkins, have distinct value, all claims as to predictive accuracy must be stringently verified by utility analyses. In essence, this means that in any prediction a chance level of accuracy must be defined, based on the observed frequency of the criterion, and then the diagnostic or forecasting technique must be contrasted with this chance level.

---

[1] Work on this project was supported by grants from the Committee on Research and the Institute of Social Sciences of the University of California, Berkeley. Statistical information and other assistance was generously provided by the Research Division of the California Youth Authority, Keith Griffiths, Chief of Research.

For example, if it is known that only 25% of men placed on parole will violate, then a prediction that every man will succeed will be correct 75% of the time. A method which eventuates in *differential* predictions for each parolee must therefore exceed this 75% base rate of accuracy if it is to be judged valuable. The more infrequent the outcome to be predicted, clearly, the more difficult it will be to exceed the undifferentiated "chance" prediction that no one will manifest the behavior.

In July of 1962 an entire issue of the journal of *Crime and Delinquency* was given over to a consideration of parole prediction tables, including discussions by a criminologist (Lejins), a probation specialist (Evjen), a sociologist (Glaser), and a psychologist (Grant). The general conclusion of this survey was that the experience table has distinct value, although its validity is far from perfect.

The particular experience table to be reported on in this study is the California Youth Authority "Base Expectancy" (CYA-BE) index, as constructed by Beverly (1959), and discussed by Beverly (1964), Beverly and Guttman (1962), and Gottfredson and Beverly (1962). The development of predictive methods having operational implications and/or effectiveness is part of a general program of research in the California Youth and Adult Corrections Agency (cf. Gottfredson, 1960; Gottfredson & Bonds, 1961; Grant, 1962; Mueller & Coon, 1964). In its current version, the CYA-BE index employs seven categories, as indicated in Table 1, with scores ranging from 0 to 68.

TABLE 1

California Youth Authority Base Expectancy Table, for Predicting Success on Parole

Variable	Weight
1. Offense	
Crimes against persons	9
Crimes against property	0
2. County of commitment	
Group 1	10
Group 2	5
Group 3	0
3. Age at first admission, to last birthday	
16 or older	12
15	6
14 or younger	0
4. Age at release to parole, to nearest birthday	
18 or older	10
17	4
16 or younger	0
5. Court of most recent commitment	
Juvenile	6
Adult	0
6. Admission status	
First admission, or return with new commitment	8
Return without new commitment	0
7. Record prior to California Youth Authority commitment	
No prior commitment and no more than two prior delinquent contacts	13
No prior commitment and three or more prior delinquent contacts	5
Prior commitment	0

For 5,132 boys released on parole during 1960, an earlier nine-category version of the BE index gave a mean score of 51.18, standard deviation 22.53. For the 2,702 who successfully completed the year on parole, the mean was 55.20, standard deviation 22.55, whereas for the 2,430 failures the values were 43.16 and 20.29. The difference of 12.04 between the two means is highly significant ($p < .001$).

Psychological Studies

Although the prediction of outcomes on parole would seem to be a natural domain for psychological measurement, most of the work, as implied above, has been done by sociologists and criminologists, using experience table methodology. Some studies have considered clinical and intuitive forecasts (see Gough, 1962; Meehl, 1954), but usually with disappointing results (cf. Glaser, 1955).

An attempt at psychological measurement was made by Clark in 1948. He constructed a 24-item recidivism scale from the MMPI (Hathaway & McKinley, 1943), contrasting the inventory responses of 45 soldiers who had been AWOL (absent without leave) once and 55 soldiers with two or more such offenses. Several of the standard MMPI scales, particularly *Pd* and *Ma*, showed slight differences between the two samples, but best results were obtained from the item analysis.

This 24-item scale was cross-validated (Clark, 1953) on a new sample of 104, consisting of 30 men AWOL only once and 74 men with two or more absences; the difference between the means on the 24-item scale for the two groups was significant at the .05 level of probability ($t = 2.02$). However, a prior effort to cross-validate Clark's scale (Freeman & Mason, 1952) produced insignificant results, and new item analyses were also unsuccessful.

A similar study was conducted by Panton (1962) who contrasted 41 violators with 41 nonviolators of parole. An item analysis of the MMPI permitted identification of 26 differentiating items. A parole violation scale (PaV) composed of these 26 items distinguished very well in the two original samples, using a cutting score of 11. On a new sample of 28 violators, 22 had scores of 11 or more, whereas on an unselected sample of 57 new parolees only 21 had scores of 11 or more.

Panton did not use Clark's recidivism scale, nor did he relate his analyses to those of Clark. A check of the items in the two scales reveals that only three appear in both: "I sometimes keep on at a thing until others lose their patience with me," "Some people are so bossy that I feel like doing the opposite of what they request, even though I know they are right," and "I have used alcohol excessively," all scored for "true." Panton's scale has a preponderance of "true" items, 23 to 3, whereas Clark's scale is more evenly balanced, 14 trues and 10 falses.

Lauber and Dahlstrom (reported in Hathaway & Monachesi, 1953) studied 35 delinquent girls in a rehabilitation program, contrasting 18 for whom treatment was successful with 17 for whom it was not. Three of the 12 scales on the MMPI revealed significant differences: the failure group was higher on *Sc* (schizophrenia), *Pt* (psychasthenia), and *Hy* (hysteria). Although *Pd* (psychopathic deviate) was the highest standard score for both groups, it did not differentiate between them.

Hathaway, Hastings, Capwell, and Bell (also reported in Hathaway & Monachesi, 1953) carried out a profile analysis for 50 delinquent girls, 17 of whom on follow-up had continued to exhibit problems of asocial behavior. Five clinical psychologists, familiar with MMPI profile interpretation, were given the set of 50 profiles and asked to classify them into groups of 33 and 17 according to the interpreter's guess as to outcome. A chance level of accuracy would be 33 "hits," or 55.2%. The actual forecasts for the individual interpreters varied from 64% to 68% correct. A composite forecast, using three out of five interpretations to determine overall classification, was correct for 35 girls, or 70% of the time.

The third instrument to be considered in this research, the CPI (Gough, 1957), has not been as extensively used in studies of parole outcomes. Perhaps the most important prior evidence on the CPI comes from the results of the California Youth Authority and Department of Corrections Pilot Intensive Counseling Organization Project (1956, 1958). In these studies it was found that the So (socialization) scale (Gough, 1960; Gough & Peterson, 1952) of the inventory differentiated among 406 parolees classified according to early violation, late violation, and no violation within a 12-month period.

Although the CPI has not been used widely in studies of parole outcomes, it has been extensively applied in studies of delinquency and asocial behavior, with very encouraging results. In particular, the So (socialization) scale (Gough, 1960; Gough & Peterson, 1952), which rests on an interactional or role-taking theory of delinquent behavior (Gough, 1948), has proved to be valid for forecasting asocial behavior among children (Scarpitti, Murray, Dinitz, & Reckless, 1960), incidence of delinquency among military personnel (Datel, 1962), rate of delinquency among military personnel (Knapp, 1963), severity of asocial behavior (Gough, 1960), delinquency in other cultures (Gough & Sandhu, 1964), and even such specific criteria as cheating on course examinations in college (Hetherington & Feldman, 1964).

In the California Youth Authority program, both the MMPI and the CPI are in widespread use, and the BE index is computed for all wards. All three instruments were therefore available for the study to be described below, and all are clearly relevant to its concerns.

METHOD AND PROCEDURE

The research sample was composed of 739 California Youth Authority wards, including all parolees with CPI and MMPI protocols from Deuel Vocational Institution and the Correctional Training Facility, Northern Branch, during 1960. A random selection of 444 subjects was made for initial study, reserving 295 for cross-validation.

Determination of parole outcome was made in April of 1963 when all cases were classified as successes or failures. Failure was defined as revocation of parole, either for violation of terms or for commitment of a new offense. The minimum period on parole for any ward classified as successful was 2 years, and many had completed $2\frac{1}{2}$ or 3. In the initial sample there were 183 violators and 261 nonviolators, and in the cross-validating sample there were 130 violators and 165 nonviolators.

For all 739 wards, initial admission was to the Deuel Vocational Institution, where processing, orientation, and testing is conducted at a Reception-Guidance Center. It is believed that this period of guidance and adaptation enhances the dependability and validity of the psychological testing results, although its primary goal is of course humanitarian and therapeutic. Assignments for work and training are planned during this period, and every effort is made to enlist the full participation of the ward in making these decisions. Extensive diagnostic study of each inmate is also conducted, including the administration of the CPI and MMPI, tests of ability and educational achievement, and life-history interviews.

Multiple-regression technique was the principal method of analysis utilized in this study.[2] Six equations were developed on the initial sample, and then verified on the cross-validating sample. The equations were derived from the following clusters of diagnostic data: (*a*) Base Expectancy (BE) and CPI; (*b*) BE and MMPI; (*c*) BE, CPI, and MMPI; (*d*) CPI alone; (*e*) MMPI alone; and (*f*) CPI and MMPI. The equations were compared with each other, and also with BE alone, and with the earlier scales for recidivism and parole violation developed by Clark and Panton.

FINDINGS

Descriptive data for all four subsamples are presented in Table 2. Violators do not differ from nonviolators on age, but clearly significant differences for both the initial and cross-validating samples emerge on the BE index ($p < .01$).

On the CPI, the So (socialization) scale differentiates significantly ($p < .01$) between violators and nonviolators in both samples, and the Sc (self-control) differentiates at the .01 level in the first sample and the .05 level in the second.

On the MMPI, Welsh's scales for anxiety (A) and repression (R) and Barron's scale for ego strength (Es) were scored in addition to the 13 standard measures. The only variable of these 16 which differentiated between violators and nonviolators was the *K*-corrected *Ma* (hypomania) scale ($p < .05$ in both comparisons).

All values for the CPI and MMPI in Table 2 are in raw score form. For descriptive purposes, raw score means for the initial sample were converted to standard scores. Three CPI scales revealed mean standard scores of approximately 30 or below: responsibility, socialization, and intellectual efficiency. On the MMPI, *Pd* (psychopathic deviate) was highest with a mean standard score of approximately 77, followed by *Sc* (schizophrenia) at 72 and *D* (depression) at 69.

As indicated earlier, the principal analytic technique was the derivation of multiple-regression equations in the original sample, seeking to maximize the prediction of the success versus failure dichotomy. The six equations so developed are presented below:

$$\text{BE} + \text{CPI, Success} = 21.328 - .411\text{Sp} + .308\text{So} + .236\text{Sc} + .435\text{Cm} + .444\text{BE} \qquad [1]$$

$$\text{BE} + \text{MMPI, Success} = 44.585 - .159Mf - .185Pa - .326Ma + .493\text{BE} \qquad [2]$$

---

[2] This work was conducted at the Computer Center of the University of California, Berkeley. The writers wish to thank the Center for time granted for this study, and also Quintin Welch, senior statistician at the Institute of Personality Assessment and Research for conducting the statistical analyses.

TABLE 2

Descriptive Statistics for the Initial and Cross-Validating Samples on the Variables Indicated

Variables	Initial sample					Cross-validating sample				
	Violators $N = 183$		Nonviolators $N = 261$		*diff*	Violators $N = 130$		Nonviolators $N = 165$		*diff*
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	
Age at release	20.18	1.32	20.23	1.12	−0.05	20.12	1.03	20.33	1.12	−0.21
Base Expectancy	36.08	7.88	40.29	9.14	−4.21**	36.06	7.88	40.68	9.35	−4.62**
CPI										
Do	21.15	6.08	20.29	6.01	0.86	20.67	5.80	20.61	5.95	0.06
Cs	15.05	4.55	14.64	4.24	0.41	15.38	4.45	14.46	4.49	0.92
Sy	21.38	5.56	20.67	5.56	0.71	20.93	5.04	20.90	5.06	0.03
Sp	32.76	6.04	31.28	6.02	1.48*	32.20	5.76	31.59	5.90	0.61
Sa	19.41	4.24	18.57	4.24	0.84*	19.42	3.92	18.75	4.05	0.67
Wb	31.98	5.96	32.42	6.12	−0.44	31.52	6.70	31.67	6.47	−0.15
Re	20.14	5.61	21.14	5.70	−1.00	20.00	5.95	20.61	5.83	−0.61
So	25.27	4.73	26.93	5.26	−1.66**	24.77	4.90	26.84	5.59	−2.07**
Sc	23.81	7.50	26.14	7.35	−2.33**	23.77	8.05	25.61	7.41	−1.84*
To	14.68	5.12	15.51	5.29	−0.83	14.77	5.90	15.07	5.31	−0.30
Gi	16.01	6.35	16.87	5.81	−0.86	16.42	6.23	17.05	5.88	−0.63
Cm	24.27	3.22	24.58	3.17	−0.31	23.74	3.79	24.55	3.23	−0.81*
Ac	20.13	5.30	20.93	5.28	−0.80	19.57	5.66	20.64	5.08	−1.07
Ai	13.38	4.15	13.92	4.27	−0.54	13.39	4.62	13.19	4.74	0.20
Ie	30.56	6.95	31.03	6.42	−0.47	30.48	6.80	30.12	6.47	0.36
Py	8.88	3.01	8.83	2.69	0.05	8.68	2.86	8.92	3.30	−0.24
Fx	8.47	3.30	8.24	3.28	0.23	8.51	3.75	7.79	3.45	0.72
Fe	16.25	3.70	16.32	3.58	−0.07	16.40	3.52	16.16	3.63	0.24
MMPI										
L	4.64	2.82	4.89	2.56	−0.25	4.91	2.82	4.90	2.45	0.01
F	10.45	6.13	9.48	6.56	0.97	11.25	7.36	9.72	6.32	1.53
K	12.86	4.93	13.30	4.58	−0.44	12.92	5.64	12.67	4.79	0.25
Hs + .5*K*	14.50	4.46	14.76	5.00	−0.26	14.65	5.05	14.13	4.47	0.52
D	24.10	5.23	24.35	5.80	−0.25	23.58	5.01	23.26	5.11	0.32
Hy	20.86	5.48	21.00	6.10	−0.14	21.24	5.37	20.26	5.81	0.98
Pd + .4*K*	30.55	4.07	30.33	4.44	0.22	30.80	4.71	29.47	5.03	1.33
Mf	23.95	4.49	23.47	4.94	0.48	23.77	4.95	22.98	4.97	0.79
Pa	11.83	4.20	11.31	4.27	0.52	12.00	4.16	11.61	4.33	0.39
Pt + *K*	30.54	6.35	30.76	6.60	−0.22	30.88	6.51	30.36	5.92	0.52
Sc + *K*	33.93	8.23	33.00	8.59	0.93	34.85	8.35	33.05	8.01	1.80
Ma + .2*K*	24.23	4.28	23.18	4.49	1.05*	24.44	4.92	23.00	5.02	1.44*
Si	28.83	8.37	29.06	8.11	−0.23	29.32	8.34	28.82	6.77	0.50
A	17.27	8.79	17.03	8.14	0.24	17.82	8.60	17.45	8.26	0.37
R	15.25	4.02	15.46	4.19	−0.21	14.98	4.68	15.13	3.71	−0.15
ES	42.76	5.96	42.88	5.62	−0.12	42.85	5.56	43.11	5.69	−0.26

* $p \leq .05$.
** $p \leq .01$.

$$\text{BE + CPI + MMPI, Success} = 23.409 + .281\text{CPI-So} + .254\text{CPI-Sc} - .196\text{MMPI-}Mf - .044\text{MMPI-}Pa + .466\text{BE} \qquad [3]$$

$$\text{CPI, Success} = 45.078 - .353\text{Sp} - .182\text{Sa} + .532\text{So} + .224\text{Sc} \qquad [4]$$

$$\text{MMPI, Success} = 66.363 - .081F + .065K - .055Pd - .168Mf - .456Ma \qquad [5]$$

$$\text{CPI + MMPI, Success} = 53.310 - .464\text{CPI-Sp} + .449\text{CPI-So} + .186\text{CPI-Sc} - .160\text{MMPI-}F - .144\text{MMPI-}Ma \qquad [6]$$

The multiplying weights are for use with raw scores (*K*-corrected raw scores on the MMPI), and the constants are adjusted so that the mean value for an unselected sample of delinquents should approach 50.00. In

TABLE 3

COMPARISON OF VIOLATORS AND NONVIOLATORS FROM THE ORIGINAL SAMPLE, ON THE PREDICTORS INDICATED

Predictors	Violators $N = 183$		Nonviolators $N = 261$		*diff*	*t*
	M	*SD*	*M*	*SD*		
Base Expectancy	36.08	7.88	40.29	9.14	−4.21	5.06**
Special scales						
Recidivism (Clark)	12.60	2.48	12.11	2.66	0.49	1.99*
Parole violation (Panton)	12.79	4.02	12.31	3.56	0.48	1.30
Equations						
BE + CPI	47.85	5.27	51.52	5.99	−3.67	6.66**
BE + MMPI	48.43	4.49	51.07	5.17	−2.64	5.57**
BE + CPI + MMPI	48.16	4.74	51.30	5.65	−3.14	6.14**
CPI	48.75	4.41	50.84	4.45	−2.09	4.89**
MMPI	49.57	2.45	50.26	2.63	−0.69	2.80**
CPI + MMPI	48.74	4.59	50.89	4.37	−2.15	5.00**

* $p < .05$.
** $p \leq .05$.

deriving the equations, *F* values for the inclusion or rejection of variables were set at the .05 level; an arbitrary decision was also made to limit the number of variables to four or five for each equation.

The effectiveness of the six equations in the initial sample may be judged from the data in Table 3. Table 3 also includes findings for the BE index alone, and for the two special scales from the MMPI.

All six equations, as would be expected, differentiate significantly between violators and nonviolators, and three of them exceed the level of significance given by BE alone. A fourth, that based on the two inventories, is almost equal to the BE index. The two special scales from the MMPI do not fare as well. Panton's scale for parole violation gives only a slight and insignificant difference between the two samples, and the proposed cutting score of 11 would be quite ineffective. Clark's recidivism scale yields a difference just significant at the .05 level, and as a quick diagnostic measure may therefore have value.

Cross-validation of the six equations is a necessity, as differentiation within the initial sample where the equations were developed is insufficient evidence of utility. Table 4 offers this cross-validational evidence, and permits a comparison of the relative predictive efficiency of all nine bases of forecasting considered in this study.

The BE index shows a mean difference of − 4.63, slightly greater than in the original sample, but because of the smaller number of cases the *t* ratio drops from 5.06 to 4.52; this value, nonetheless, is significant well beyond the .01 level of probability. Clark's recidivism scale just fails to achieve significance at the .05 level, and Panton's parole violation scale again gives an insignificant difference.

The largest *t* ratio is provided by the five-variable regression equation drawn from the CPI and the BE index. All six of the equations yield significant differences ($p < .01$), but for only three do the *t* ratios exceed that given by BE alone. It is worth noting that the combination of BE + CPI gives a slightly larger *t* ratio than that yielded by BE, the CPI, and the MMPI.

Having found that the equations possess differentiating validity, and that the discrimination afforded by BE can be augmented by the addition of personality inventory data, the next question becomes one of the degree of accuracy which might be achieved through use of these equations. From the standpoint of utility, we wish to see if any of the equations can surpass the accuracy of an undifferentiated prediction that all parolees will succeed; this "chance" prediction would be correct for 165 subjects, that is, 56% of the time.

TABLE 4

Comparison of Violators and Nonviolators from the Cross-validating Sample, on the Predictors Indicated

Predictors	Violators $N = 130$		Nonviolators $N = 165$		*diff*	*t*
	M	*SD*	*M*	*SD*		
Base Expectancy	36.05	7.88	40.68	9.35	−4.63	4.52**
Special scales						
Recidivism (Clark)	12.87	2.35	12.32	2.48	0.55	1.95
Parole Violation (Panton)	13.10	4.48	12.76	2.75	0.34	0.76
Equations						
BE + CPI	47.70	5.40	51.38	6.27	−3.68	5.41**
BE + MMPI	48.40	4.75	51.35	5.69	−2.95	4.85**
BE + CPI + MMPI	48.02	5.14	51.37	5.97	−3.35	5.17**
CPI	48.72	4.40	50.52	4.81	−1.80	3.35**
MMPI	49.48	3.03	50.41	3.03	−0.93	2.62**
CPI + MMPI	48.62	4.83	50.56	4.90	−1.94	3.40**

* $p \leq .05$.
** $p \leq .01$.

To avoid capitalizing on peculiarities in the distributions of scores, a cutting point was set on each of them so that 130 subjects would fall below the point (or as close to 130 as possible), and 165 above. For the four equations involving more than one instrument, this cutting point was set at 48 and below (all such cases designated as failures); for the MMPI the cutting point for failures was set at 49 and below, and for the CPI alone the point was set at 48 and below. BE alone was also evaluated, using a cutting score of 35 and below.

Using these cutting points, the percentage of "hits" for BE alone and for each equation was as follows: BE alone, 59%; BE + CPI, 63%; BE + MMPI, 60%; BE + CPI + MMPI, 63%; CPI alone, 60%; MMPI alone, 55%; and CPI + MMPI, 60%. BE alone correctly classified 175 men, 75 violators and 100 nonviolators. The best equation (BE + CPI) correctly classified 186 parolees, 76 of the 130 violators and 110 of the 165 nonviolators. The second best equation (BE + CPI + MMPI) correctly classified 185 of the 295 cases, and the third best (BE + MMPI) correctly identified 178 of the 295 parolees.

These accuracy figures are not overly impressive, even if statistically significant. However, if the typical number of boys going on parole per year in California is 5,000, as suggested by the analysis of Gottfredson and Beverly (1962), this would mean that the BE + CPI equation would be correct in 3,150 instances, whereas the chance level of accuracy in forecasting would be 2,800 hits. BE alone would be accurate 2,950 times.

Psychological Meaning

Our interest now turns to a consideration of the psychological meaning of higher and lower scores on the equations.[3] Put as a question, we are asking "What sort of person is it, in everyday language and description, who attains a high or a low score on this equation?" To secure data relevant to this question, three of the equations (CPI alone, MMPI alone, and CPI + MMPI) were applied to a sample of 100 adult males who had been studied in intensive assessment at the University of California Institute of Personality Assessment and Research. The equations involving the BE index, unfortunately, could not be used in this analysis, as data for use in computing BE were not available.

Each of the 100 men had been observed and studied individually and in groups over a period of six days, and had been described by a panel of 10 observers using the Gough ACL (Gough & Heilbrun, 1965). This List contains 300 words, drawn from a context of everyday usage. When employed by a panel of observers a point is earned for a word each time it is checked by an observer. With 10 observers this means that such a score can vary from 0 to 10. The 300 adjectival or descriptive scores thus generated may then be correlated with inventory data, regression scores, or any other variable whose interpersonal implications one wishes to discover.

---

[3] Cf. Gough (1965) for a discussion of this conception of psychological meaning.

The 100 assessment males had a mean score of 52.77, standard deviation 4.47, on the equation composed of the social presence, self-acceptance, socialization, and self-control scales of the CPI. The 10 adjectives from the observers' descriptions having highest correlations with the CPI equation have been chosen for listing, and also the 10 adjectives having lowest correlations. The 10 words most descriptive of high scorers and the correlations, were these:

conscientious (.35)
conservative (.38)
conventional (.34)
mild (.32)
moderate (.46)
modest (.37)
painstaking (.41)
patient (.33)
quiet (.32)
unassuming (.35)

The 10 words most descriptive of men with low scores on the equation were:

adventurous (−.41)
egotistical (−.39)
mischievous (−.44)
opinionated (−.37)
outspoken (−.39)
pleasure-seeking (−.46)
quick (−.38)
shrewd (−.41)
touchy (−.38)
witty (−.39)

As far as was known, no one of these 100 men was delinquent. Nevertheless, within this sample the CPI equation does appear to define a continuum of personality; furthermore, this continuum has clear relevance to the ability to live by stringently specified rules, such as would obtain on parole. The CPI equation, developed to forecast success on parole and shown in cross-validation to be capable of making such forecasts, thus reveals a clear pattern of psychological meaning when studied within the normal, nondelinquent range.

The MMPI equation, including the *F*, *K*, psychopathic deviate, masculinity-femininity, and paranoia scales, gave a mean of 50.16, standard deviation of 2.59, when applied to the 100 men. The 10 adjectives having highest positive correlations with the MMPI equation were these:

conscientious (.17)
conservative (.15)
dependable (.16)
modest (.20)
sincere (.15)
trusting (.17)
unaffected (.25)
unambitious (.15)
unassuming (.22)
wholesome (.16)

The 10 words with lowest correlations were:

distrustful (−.21)
imaginative (−.21)
irritable (−.20)
nervous (−.25)
prejudiced (−.21)
resentful (−.23)
sensitive (−.21)
touchy (−.23)
unfriendly (−.20)
wary (−.27)

The pattern is slightly different from that observed for the CPI equation, but the implications still seem clear for a readier adaptation to convention at the high end and greater likelihood of waywardness at the low.

The final equation to be considered in this adjectival analysis is that based on the CPI and the MMPI together. The five scales included were the social presence, socialization, and self-control scales from the CPI, and the *F* and hypomania scales from the MMPI. This CPI + MMPI equation yielded a mean of 50.41, standard deviation 4.63, for the 100 men. The 10 adjectives having highest positive correlations with the scores on the equation were these:

conscientious (.34)
conservative (.38)
conventional (.33)
mild (.32)
moderate (.45)
modest (.38)
painstaking (.39)
patient (.32)
quiet (.32)
unassuming (.36)

The 10 items with lowest correlations were these:

adventurous (−.39)
egotistical (−.39)
mischievous (−.42)
outspoken (−.36)
pleasure-seeking (−.46)
quick (−.37)
restless (−.36)
shrewd (−.37)
touchy (−.38)
witty (−.37)

The two lists are identical to those cited above for the CPI alone, with the exception of restless in place of opinionated. We may therefore conclude that the CPI + MMPI equation carries essentially the same psychodynamic implications as those delineated by the equation based solely on the CPI.

Although the equations combining the inventory scales with the BE index could not be computed, it seems plausible to suppose that their implications would be very similar to those of the three equations which have been analyzed. One might say that the parolee predicted to be successful on the basis of these equations is a person who is conscientious and moderate, not in any way flamboyant and perhaps even unduly subdued. His opposite, for whom negative forecasts will be made, is a more narcissistic, restless, undercontrolled individual, too sure of himself and too quick to take offense when blocked or criticized.

DISCUSSION

As shown in Tables 2 and 3, the best prediction of parole outcome from a single source is that given by the BE index. This may constitute another illustration of the axiom that the best predictor of future behavior is past behavior. The weightings in the index, it is apparent, do stress such past behavior. The earlier a boy's first commitment, the more extensive his prior record of delinquency, and the degree to which his offense of commitment involves property, the poorer will be his BE forecast for success on parole. This forecast will also be more guarded if he comes from an urban environment, is younger at time of release, has been treated in an adult rather than a juvenile court, and if he is a previous parole violator.

The "problem" with the BE technique, from the psychological standpoint, is that it is a fixed quantity. A ward might be admitted and undergo significant therapeutic change, to such an extent that his parole prospects become excellent. Yet, if his file at the time of entry leads to a low BE score, he will continue to be designated a poor risk by the BE method, no matter how much he has changed.

The advantage of prediction from inventory data is that current status is evaluated. The disadvantage is that the accuracy of prediction is, at least in our data, inferior to that of the life history material. A combination of the two sources of information, however, appears to be superior to either taken alone.

The evidence reported in this paper represents only a first step toward optimum utilization of these two sources of information. Should one simply combine the two kinds of scores into a single equation, as we have done, or should one search for cases with sharp discrepancies between the BE and inventory forecast, and then determine new procedures of forecasting for them? This and many other questions await study, and we hope that other investigators will join us in conducting research on these issues.

Attention should also be paid to the kinds of inventory measures which revealed greatest promise in our analyses. From the CPI, the major contributors were those scales having to do with the internalization of value and the management of impulse. From the MMPI, the major contributors were measures of modality and strength of impulse. Indices of psychopathology, as customarily conceived, were less predictive and the syndrome defined by the personality equations was more interpersonal and interactional than psychopathological and morbid. That is to say, good and poor risks as indicated by the equations seem to be two different kinds of people, neither abnormal, but having differing capacities for adaptation to a determined and strictly regulated regime.

REFERENCES

Barron, A. J. An experiment with Ohlin's prediction report. *Crime and Delinquency,* 1962, **8**, 276–281.

Beverly, R. F. A method of determination of base expectancies for use in the assessment of effectiveness of correctional treatment. Research Report No. 3, February 1959, Sacramento, California, Department of the Youth Authority.

Beverly, R. F. Base expectancies and the initial home visit research schedule. Research Report No. 37, January 1964, Sacramento, California, Department of the Youth Authority.

Beverly, R. F., & Guttman, Evelyn S. An analysis of parole performance by institution of release. Research Report No. 31, November 1962, Sacramento, California, Department of the Youth Authority.

Burgess, E. W. Factors determining success or failure on parole. In A. A. Bruce (Ed.), *The workings of the indeterminate sentence law and the parole system in Illinois.* Springfield, Ill.: Illinois State Board of Parole, 1928. Pp. 205–249.

California Youth Authority and State Department of Corrections. Technical Report Series, 1956, 1958, Pilot Intensive Counseling Organization Project, Deuel Vocational Institution, Tracy, California.

Clark, J. H. Application of the MMPI in differentiating AWOL recidivists from non-recidivists. *Journal of Psychology,* 1948, **26**, 229–234.

CLARK, J. H. Additional applications of the AWOL recidivist scale. *Journal of Clinical Psychology,* 1953, **9**, 62–64.

DATEL, W. E. Socialization scale norms on military samples. *Military Medicine,* 1962, **127**, 740–744.

EVJEN, V. H. Current thinking on parole prediction tables. *Crime and Delinquency,* 1962, **8**, 215–238.

FREEMAN, R. A., & MASON, H. M. Construction of a key to determine recidivists from non-recidivists using the MMPI. *Journal of Clinical Psychology,* 1952, **8**, 207–208.

GILLIN, J. L. Parole prediction in Wisconsin. *Sociology and Social Research,* 1950, **34**, 407–414.

GLASER, D. The efficacy of alternative approaches to parole prediction. *American Sociological Review,* 1955, **20**, 283–287.

GLASER, D. Prediction tables as accounting devices for judges and parole boards. *Crime and Delinquency,* 1962, **8**, 239–258.

GLUECK, S., & GLUECK, ELEANOR T. *500 criminal careers.* New York: Knopf, 1930.

GLUECK, S., & GLUECK, ELEANOR T. *Unraveling juvenile delinquency.* New York: The Commonwealth Fund, 1950.

GLUECK, S., & GLUECK, ELEANOR T. Early detection of future delinquents. *Journal of Criminal Law, Criminology, and Police Science,* 1956, **47**, 174–182.

GOTTFREDSON, D. M. Efficiency of two methods of prediction of parole success. *American Psychologist,* 1960, **15**, 580. (reference)

GOTTFREDSON, D. M., & BEVERLY, R. F. Development and operational use of predictive methods in correctional work. In Edwin Goldfield (Ed.), *Proceedings of the Social Statistics Section of the American Statistical Association.* Washington, D. C.: ASA, 1962. Pp. 54–61.

GOTTFREDSON, D. M., & BONDS, J. A. *A manual for intake base expectancy scoring (Form CDC-BE 61A).* Sacramento, Calif.: Research Bulletin, California State Department of Corrections, April 1961.

GOUGH, H. G. A sociological theory of psychopathy. *American Journal of Sociology,* 1948, **53**, 359–366.

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

GOUGH, H. G. Theory and measurement of socialization. *Journal of Consulting Psychology,* 1960, **24**, 23–30.

GOUGH, H. G. Clinical versus statistical prediction in psychology. In L. J. Postman (Ed.), *Psychology in the making.* New York: Knopf, 1962. Pp. 526–584.

GOUGH, H. G. The conceptual analysis of psychological test scores and other variables. *Journal of Abnormal Psychology,* 1965, **70**, 294–302.

GOUGH, H. G., & HEILBRUN, A. B., JR. *The Adjective Check List Manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1965.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology,* 1952, **16**, 207–212.

GOUGH, H. G., & SANDHU, H. S. Validation of the CPI socialization scale in India. *Journal of Abnormal and Social Psychology,* 1964, **68**, 544–547.

GRANT, J. D. It's time to start counting. *Crime and Delinquency,* 1962, **8**, 259–264.

HART, H. Predicting parole success. *Journal of Criminal Law and Criminology,* 1923, **14**, 405–413.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: The Psychological Corporation, 1943.

HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. Minnesota Press, 1953.

HETHERINGTON, E. MAVIS, & FELDMAN, S. E. College cheating as a function of subject and situational variables. *Journal of Educational Psychology,* 1964, **55**, 212–218.

KNAPP, R. R. Personality correlates of delinquency rate in a Navy sample. *Journal of Applied Psychology,* 1963, **47**, 68–71.

LEJINS, P. P. Parole prediction—an introductory statement. *Crime and Delinquency,* 1962, **8**, 209–214.

MANNHEIM, H., & WILKINS, L. T. *Prediction methods in relation to Borstal training.* London: Her Majesty's Stationery Office, 1955.

MEEHL, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. Minnesota Press, 1954.

MONACHESI, E. D. American studies in the prediction of recidivism. *Journal of Criminal Law and Criminology,* 1950, **41**, 268-289.

MUELLER, P. F. C., & COON, DOROTHY. Parole outcome prediction for male opiate users in the Los Angeles area Narcotic Treatment Control Program. Research Report No. 18, June 1964, Sacramento, Calif., Research Division, Department of Corrections, Youth and Adult Corrections Agency.

OHLIN, L. E. *Selection for parole.* New York: Russell Sage Foundation, 1951.

PANTON, J. H. Use of the MMPI as an index to successful parole. *Journal of Criminal Law, Criminology, and Police Science,* 1962, **53**, 484-488.

SCARPITTI, F. R., MURRAY, ELLEN, DINITZ, S., & RECKLESS, W. C. The "good" boy in a high delinquency area: Four years later. *American Sociological Review,* 1960, **25**, 555–558.

SCHUESSLER, K. F. Parole prediction: Its history and status. *Journal of Criminal Law, Criminology, and Police Science,* 1954, **45**, 425–431.

VOLD, G. B. *Prediction methods and parole.* Hanover, N. H.: The Sociological Press, 1931.

(Received January 18, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 442–445

TEMPORAL GRADIENT OF FEAR AROUSAL IN PSYCHOPATHS

ROBERT D. HARE[1]
University of British Columbia

Skin conductance of psychopathic (P) and nonpsychopathic (NP) criminals and noncriminal controls (N = 11 each group) was monitored while the numbers 1–12 were consecutively presented on a memory drum. After the 1st trial, Ss were informed that on each of the remaining 5 trials they would receive a strong electric shock when Number 8 appeared. Log conductance increases in the interval prior to shock were significantly smaller and began later for the P than for the NP Ss. The psychopaths were less responsive to shock than were the NP criminals but did not differ significantly from the noncriminal controls. There were no significant differences among groups in the rate of recovery from the effects of shock. The results were consistent with the hypothesis that the temporal gradient of fear arousal and response inhibition is steeper for psychopaths than for normal persons.

Compared to normal persons, the behavior of the psychopath[2] appears to be guided more by immediate events than by those relatively remote in time (Cleckley, 1959; Karpman, 1961). For instance, the satisfaction of his immediate needs is often obtained at the risk of receiving some future and possibly severe, punishment. The author (Hare, 1965b) has recently proposed that this and other aspects of psychopathic behavior can be accounted for by postulating that the temporal gradient of fear arousal and response inhibition is steeper for psychopaths than for normal persons. This is based upon, among other things, the psychopath's relative lack of anxiety (Cleckley, 1959; Lykken, 1957), and his difficulty in developing and retaining conditioned fear responses (Hare, 1965c; Lykken, 1957). Thus, it is assumed that as the temporal remoteness of anticipated punishment increases, the amount of fear elicited by cues associated with the punishment decreases (cf., Miller, 1959; Renner, 1964), with the rate of decrease being greater for psychopathic than for nonpsychopathic persons. To the extent that fear mediates response inhibition (Mowrer, 1960), the psychopath would therefore be less likely than would the normal person to inhibit responses for which anticipated punishment is remote in time. Consistent with this are the results of a recent study (Hare, 1965a) in which skin conductance increases to an impending electric shock were found to be related to scores on the Psychopathic Deviate (*Pd*) scale of the MMPI. As the administration of electric shock became imminent, subjects with high *Pd* scores began to react somewhat later, more slowly, and to a lesser degree than did those with low *Pd* scores.

The present study was similar to the previous one except that subjects were selected on the basis of clinical diagnosis rather than on the basis of inventory scores.

METHOD

Subjects were all male volunteers and consisted of psychopathic (Group P) and nonpsychopathic (Group NP) inmates of the British Columbia Penitentiary, and students (Group C) obtained from a local adult education center. Selection of the psychopaths was made by staff psychologists on the basis of a 12-item checklist of criteria derived from Cleckley (1959). In addition to satisfying the requirements of the checklist, most of the subjects in Group P had been classified as psychopathic at some point in their criminal history.

The characteristics of each group are presented in Table 1. None of the differences among groups were significant. Most of the subjects in each group were from the lower socioeconomic levels.

Skin resistance was measured with a Lafayette GSR recorder and ⅜-inch diameter zinc-zinc sulphate electrodes (Lykken, 1959), taped to the finger-print area of the first and third fingers of the right hand. A Psychological Instruments Stimulator was used to deliver a brief electric shock through finger electrodes attached to the first and third fingers of the left hand. The stimulus materials, adapted from Deane (1964), consisted of the Numbers 1–12 presented consecutively via a memory drum.

---

[1] Supported by a grant from the Dean's Committee on Research, U.B.C. Appreciation is extended to John Maloney, Warden of the B.C. Penitentiary, and to S. A. Miller, Director of Research and Special Services for the Vancouver School Board, for making subjects and facilities available.

[2] The term refers to the entity described by Cleckley (1959) and to what Karpman (1961) calls the primary psychopath.

Testing was conducted in a quiet room located in the appropriate institution. The subject was seated directly in front of the memory drum and, after the GSR and shock electrodes had been attached, was required to sit quietly for about 5 minutes. During this period the resting skin resistance was determined. The subject was then told that a few sample shocks would be administered to determine the intensity to be used during the experiment. The intensity was gradually increased until the subject indicated that he was unwilling to accept anything stronger. The mean intensities of shock arrived at by this procedure and subsequently used in the experiment were 3.8, 3.3, and 3.7 milliamperes for Groups P, NP and C, respectively. These intensities did not differ significantly from each other.

Several minutes were allowed for the subject to recover from the effects of the sample shocks. His attention was then directed to the memory drum and the following instructions given: "Through this window here you will see a series of consecutive numbers running from 1 to 12. I want you to watch the numbers closely. No shocks will be given until later." The series was presented once, and the subject was then told: "The same series will be repeated several more times, but from now on you will receive a shock everytime number eight comes up. The intensity of the shock will be the same as that which you earlier indicated was your maximum." The series was presented five times, with each number appearing for 3 seconds. About 21 seconds elapsed between successive trials. At the conclusion of the experiment the subject was informally questioned about his experiences.

RESULTS AND DISCUSSION

Skin resistance readings were converted to log conductance units prior to performing any computations. This was done in an attempt to normalize the distribution of scores and to minimize the influence of the prestimulus level upon response magnitude (see Lacey & Siegel, 1949; Schlosberg & Stanley, 1953). Unless otherwise indicated, all tests of significance reported involved analysis of variance followed in each case by Duncan's test for multiple comparisons among means (Edwards, 1960).

The mean resting log conductance of Group P was significantly lower ($p < .05$) than that of either Group NP or C. This is consistent with other studies in which psychopaths were found to have a relatively low level of autonomic tension (e.g., Claridge & Herrington, 1963; Walton, 1960). The level of log conductance during each trial was determined by taking readings 1½ seconds after the appearance of each stimulus number. This was done because the GSR activity that occurred was not always due simply to the presentation of a stimulus number. In many cases, it appeared that responses were elicited by the sound of the memory drum bringing the next number into view, or by covert cues not under experimental control. Furthermore, many subjects did not give clear-cut GSRs, but rather showed a gradual, though often somewhat irregular, change in log conductance. Preliminary inspection of the records suggested that a reasonable procedure would be to measure log conductance midway through the presentation of each stimulus number.

TABLE 1

CHARACTERISTICS OF THE THREE GROUPS

Group	*N*	Age		Education		Revised beta IQ	
		M	*SD*	*M*	*SD*	*M*	*SD*
P	11	27.4	4.1	10.5	1.7	111.8	8.8
NP	11	26.7	4.2	9.8	1.8	107.3	10.2
C	11	24.5	3.1	11.2	0.5	109.4	9.8

During Trial 1 (no shock) each group showed a gradual though insignificant decrease in log conductance. The results for Trials 2 and 6 (first and last shock trials) are plotted in Figure 1. It is apparent that the psychopathic subjects (Group P) differed from the other subjects, not only in having a generally lower level of log conductance, but also in the pattern of change as shock approached in time. Thus, for both Trials 2 and 6, the increase in log conductance during the period just prior to shock (stimulus Numbers 4 to 7) was significantly smaller for Group P than for either Group NP or C ($p < .05$ in each case). As a matter of fact, Group P showed virtually no increase in log conductance during this period.

While it was not possible to determine directly the magnitude of response to the shock itself since it was superimposed upon a background of rapidly rising conductance, some

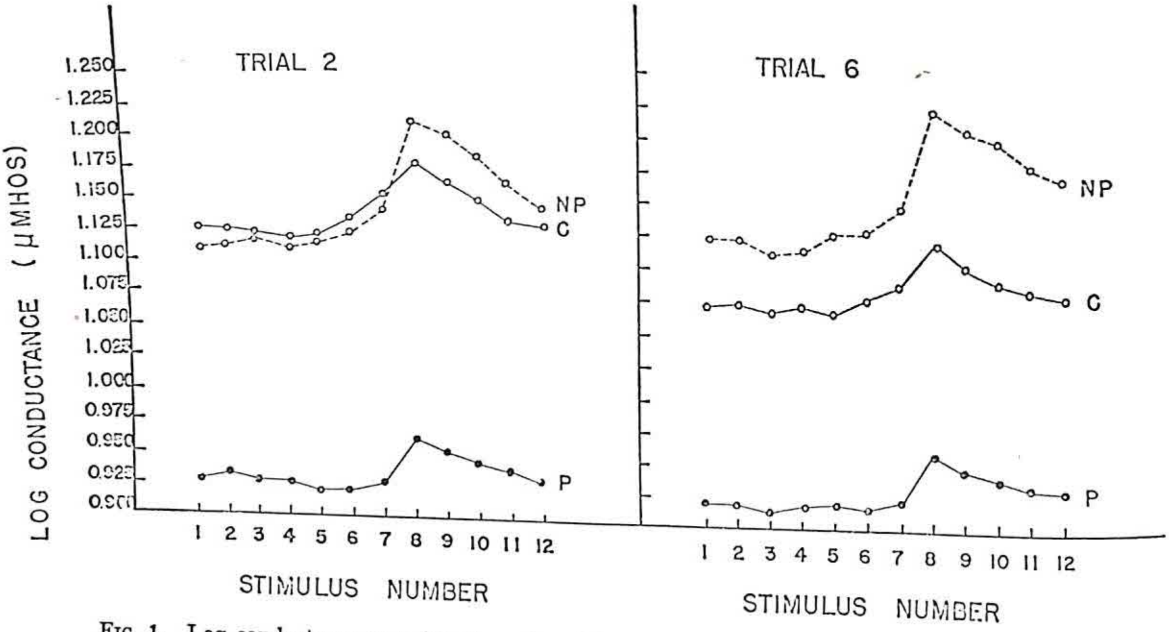


FIG. 1. Log conductance as a function of anticipated shock (shock at stimulus Number 8).

indication of the overall effects of shock and its anticipation was obtained by computing the change in log conductance from stimulus Number 1 to 8. On each trial, the only significant difference was between Groups P and NP ($p < .05$ in both cases), with the latter showing the greater increase. The magnitude of this change in log conductance from stimulus Number 1 to 8 was not significantly correlated with the level of log conductance at stimulus Number 1, either for Trial 2 ($r = .08$) or for Trial 5 ($r = .05$). It is unlikely therefore that the log conductance changes reported were influenced to any great extent by differences between groups in base line or prestimulus level.

The fact that the overall effects of shock on Group P were relatively small when compared with Group NP but not when compared with Group C provides only limited support for the commonly held hypothesis that psychopaths are less responsive to noxious stimulation than are normal persons (e.g., Tong, 1959; Walton, 1960). It is possible that what differences were observed between Groups P and NP were due to the presence of a greater proportion of highly anxious individuals in Group NP than would ordinarily be found in a nonincarcerated population. It is noteworthy that an early study by Lindner (1942) found that "constitutional psychopathic inferiors" were less responsive (GSR deflections) prior to, but more responsive during, noxious stimulation than were other criminals.[3] However, his criteria of psychopathy suggest that a considerable number of neurotic, emotionally disturbed subjects may have been included in his psychopathic group. Lindner's finding that psychopaths recovered rapidly from the effects of stimulation was not supported by the present study. An analysis of the decrease in log conductance from stimulus Numbers 8 to 12 revealed no significant differences among groups for either trial. It is possible that the lack of consistency in these studies reflects the use of different procedures and criteria of psychopathy.

The finding that only Group C showed a significant ($p < .05$) decrease in log conductance from Trial 2 to Trial 6 can be explained by the fact that Groups P and NP consisted of penitentiary inmates who were concerned about the effects their performance might have upon their prison records, and who consequently remained tense throughout the experiment.

---

[3] Schachter and Latane (1964) have recently reported that the pulse rate of psychopathic criminals was slightly (though not significantly) higher than that of nonpsychopathic criminals, both before and after an injection of epinephrine. On the basis of this and other relatively indirect evidence, the authors conclude that the psychopath is more autonomically responsive to stressful stimuli than are normal persons.

To the extent that skin conductance increases prior to shock can be considered an indicant of fear arousal, the present results suggest that relatively little fear is elicited in psychopathic persons in the interval prior to anticipated punishment. This is consistent with the hypothesis that psychopaths can be characterized by a steeper than normal gradient of fear arousal and response inhibition. Presumably, cues associated with future punishment are incapable of generating sufficient fear in the psychopath for immediate behavior to be inhibited. The psychopath's impulsivity and his repeated failure to inhibit antisocial behavior and hence to avoid punishment might be understood in these terms.

Subsequent research along these lines will involve the use of several autonomic measures, longer temporal intervals, and various types, intensities and probabilities of punishment.

REFERENCES

CLARIDGE, G. S., & HERRINGTON, P. H. Excitation-inhibitions and the theory of neurosis: A study of the sedation threshold. In H. J. Eysenck (Ed.), *Experiments with drugs*. New York: MacMillan, 1963. Pp. 131–168.

CLECKLEY, H. Psychopathic states. In S. Arieti (Ed.), *American handbook of psychiatry*. Vol. 1. New York: Basic Books, 1959. Pp. 567–588.

DEANE, G. E. Human heart rate responses during experimentally induced anxiety: A follow-up with controlled respiration. *Journal of Experimental Psychology*, 1964, **67**, 193–195.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1960.

HARE, R. D. Psychopathy, fear arousal, and anticipated pain. *Psychological Reports*, 1965, **16**, 499–502. (a)

HARE, R. D. A conflict and learning theory analysis of psychopathic behavior. *Journal of Research in Crime and Delinquency*, 1965, **2**, 12–19. (b)

HARE, R. D. Acquisition and generalization of a conditioned fear response in psychopathic and nonpsychopathic criminals. *Journal of Psychology*, 1965, **59**, 367–370. (c)

KARPMAN, B. The structure of neurosis: With special differentials between neurosis, psychosis, homosexuality, alcoholism, psychopathy, and criminality. *Archives of Criminal Psychodynamics*, 1961, **4**, 599–646.

LACEY, O. L., & SIEGEL, P. S. An analysis of the unit of measurement of the galvanic skin response. *Journal of Experimental Psychology*, 1949, **39**, 122–127.

LINDNER, R. M. Experimental studies in constitutional psychopathic inferiority. *Journal of Criminal Psychopathology*, 1942, **3**, 252–276.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, 1957, **55**, 6–10.

LYKKEN, D. T. Properties of electrodes used in electrodermal measurement. *Journal of Comparative and Physiological Psychology*, 1959, **52**, 629–634.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.

RENNER, K. E. Conflict resolution and the process of temporal integration. *Psychological Reports*, 1964, **15**, 423–438.

SCHACHTER, S., & LATANE, B. Crime, cognition, and the autonomic nervous system. In D. Levine (Ed.), *Nebraska symposium on motivation: 1964*. Lincoln: Univ. Nebraska Press, 1964. Pp. 221–273.

SCHLOSBERG, H., & STANLEY, W. C. A simple test of the normality of twenty-four distributions of electrical skin conductance. *Science*, 1953, **117**, 35–37.

TONG, J. E. Stress reactivity in relation to delinquent and psychopathic behavior. *Journal of Mental Science*, 1959, **105**, 935–956.

WALTON, D. Drug addiction and habit formation: An attempted integration. *Journal of Mental Science*, 1960, **106**, 1195–1229.

(Received January 30, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 446–450

EFFECTS OF WEIGHT INTENSITY ON DISCRIMINATION THRESHOLDS OF NORMALS AND SCHIZOPHRENICS

GERALD ROSENBAUM, FRANK FLENNING,[1] AND HAROLD ROSEN[2]

Wayne State University and Lafayette Clinic

20 male chronic schizophrenics, 12 chronic disturbed schizophrenics, and 20 normal controls were tested for upper difference limens from both a 40-gm. (light) and 400-gm. (heavy) standard weight. Weight discrimination thresholds were found to be significantly elevated as a function of severity of pathology in the schizophrenic groups and also at the lighter weight intensity. Both groups of schizophrenics showed significantly greater improvement than normals with the heavy weights. The less disturbed chronic schizophrenics were not significantly different from normals at the heavy intensity. The results support the hypothesis of a schizophrenic deficit in proprioceptive acuity and suggest that this deficit is the result of insufficient proprioceptive feedback.

Previous research on weight discrimination in normal and schizophrenic subjects (Rosenbaum, Cohen, Luby, Gottlieb, & Yelen, 1959) has shown marked elevations in threshold in chronic schizophrenics and in normals treated with the drug, Sernyl. When upper difference limens were obtained from a 44-grams standard weight, nondrugged groups of normals consistently showed discrimination thresholds of less than 4 grams, while the mean threshold was 9.48 grams for the schizophrenic group and 10.46 grams for the Sernyl-drugged normals. These results suggested that the psychological deficits which were found to be associated with both schizophrenia and Sernyl administration might be attributable to a common disturbance of proprioception, interfering with the integration of response-produced feedback. The fact that Sernyl is an anesthetizing chemical compound which in small dosages closely simulates the primary symptoms of schizophrenia (Luby, Cohen, Rosenbaum, Gottlieb & Kelley, 1959) lends additional support to this interpretation.

Other investigators have also reported findings suggesting a proprioceptive disorder in schizophrenics. Sonder (1954) and Zubin (1954) have shown that the ability to judge lifted weights was a prognostic index in schizophrenics. Plumb and Meigs (1961) found that poor vibratory sensitivity at high frequencies was related to severity of schizophrenic pathology. In addition, Rado's (1953) theory that proprioceptive disorder is an innate trait of schizotypal organization postulates that extensive proprioceptive deficit may trigger the characteristic schizophrenic thought disorder by first causing a distorted awareness of the body image.

On the other hand, Salzinger (1957) found no difference between a group of 16 acute schizophrenics and a matched group of normals in their ability to discriminate weights by the method of absolute judgment. While the schizophrenics shifted significantly more than normals toward a 900-grams anchor weight, their judgments of unanchored 200-, 250-, 300-, 350-, and 400-grams weights, as well as their shifts toward a 100-grams anchor, were all comparable to normals.

Although the discrepancy between Salzinger's results and those obtained with chronic schizophrenics may be attributable to differences in the patient groups or methods of determining weight discrimination, the fact that he employed a comparatively heavy series of unanchored weights suggests an alternative hypothesis. If impaired responsiveness to proprioceptive stimulation is characteristic of schizophrenia, then increasing response-produced stimulation by providing heavier weights could significantly improve performance.

To test this hypothesis, discrimination thresholds were obtained from two groups of schizophrenics, varying in severity of pathology, and a control group of normals at two levels of weight intensity. It was hypothesized that schizophrenics would show a marked elevation in threshold at the light-weight intensity, but would approach normal discrimination with the heavy series of weights.

---

[1] Now at Southern Illinois University.

[2] Now at Marquette University School of Medicine.

METHOD

Subjects. Twenty male chronic schizophrenics were selected from the psychiatric wards of Wayne County General Hospital on the basis of mild maintenance medication and ability to demonstrate good attention and cooperation on a brief spatial- and temporal-orientation task (DeLa Garza & Worchel, 1956). A second group of 12 chronic disturbed schizophrenic males, who had not received any drug medication for at least 6 months, were chosen from the research ward of Northville State Hospital for their ability to cooperate on the weight discrimination task. All schizophrenic patients had been hospitalized for at least 3 years. Twenty male subjects composed of volunteer prisoners, attendants, and neurotic patients, all psychiatrically screened for schizophrenic symptoms, served as normal controls. The age distributions of the three groups were: chronic schizophrenic $M = 37.4$, range 32–45; chronic disturbed schizophrenic $M = 29.2$, range 25–40 normal $M = 30.1$, range 20–47. All three groups were comparable in education. The chronic disturbed group was included in the study to ascertain whether increased weight intensity would modify discrimination thresholds in a severely disturbed group of patients.

Procedure. All subjects were tested for upper difference limens from both a 40-grams (light) and a 400-grams (heavy) standard weight. Half of the subjects in each group were tested in the light-heavy order and half in the reverse order, with a 1-week interval between tests to minimize practice and anchor effects.

Upper difference thresholds were obtained by a modified method of limits, requiring the subjects to compare the standard with heavier weights presented in descending order ranging from 60 to 42 grams in the light 40-grams series and 600 to 420 grams in the heavy 400-grams series. All subjects were blindfolded with opaque goggles and instructed to "heft" the standard stimulus and comparison weight successively with the preferred hand, designating whether the first or the second weight was heavier. The order of presentation of the standard stimulus as the first or second weight was randomized within each block of comparison trials. All weights were presented in identical cylindrical cardboard containers 3.25 inches high and 2 inches in diameter.

The first 4 trials of each test series involved either a 20- or 200-grams discrimination and served as a screening and training procedure. The subjects were verbally reinforced for correct discriminations on these trials; and those subjects achieving less than 100% correct judgments were eliminated from the experiment.

On the test series, the subjects received 8 trials with each comparison weight beginning with the heaviest (56 or 560 grams) and descending in the series by either 4- or 40-grams steps until the subject reached either the lightest comparison weight or chance discrimination. If the subject showed chance discrimination (4 or more errors out of 8) before reaching the lightest weight, he was then given a block of 8 trials on the next heavier weight (ascending by 2 or 20 grams) in an effort to "zero in" on his discrimination threshold. The subjects performing at better than chance discrimination on the 44- or 440-grams weights, received a final 8 test trials on the 42- or 420-grams weights. This procedure proved satisfactory for obtaining reasonably accurate and stable individual thresholds over a wide enough range of stimuli to encompass a very heterogeneous group of subjects. Limens were obtained in only one direction, with a minimum of trials so that the time requirements of the task would not exacerbate the attention difficulties of schizophrenic patients.

Two measures of weight discrimination were derived from this procedure, (*a*) percentage of correct discriminations and (*b*) individual Weber ratio scores. The percentage of correct discriminations were determined for each subject by determining the number of correct judgments made on the 40 trials with the 56-, 52-, 48-, 44-, and 42-grams comparison weights or with their counterparts in the heavy series. Chance discrimination values were assigned when the subjects obtained large thresholds and were not tested on the lightest weights. Individual Weber ratios were computed by extrapolating the smallest weight difference at which the subject obtained 75% correct discrimination and dividing this difference by the standard weight.

RESULTS

The major results of the experiment are presented in Table 1 which shows the performance of the three groups at both weight intensities. The data for percentage of correct discriminations and individual Weber scores were each subjected to a 3 × 2 × 2 analysis of variance, permitting independent estimates of the main effects of the 3 diagnostic groups, the 2 orders of testing, and the 2 weight intensities, as well as their interactions. All analyses were performed on the square roots of the obtained scores in order to eliminate the heterogeneity of variance occasioned by the greater variability of the schizophrenic groups. These analyses are shown for both measures in Table 2.

Percentage of correct discriminations. Inspection of Table 1 and the results of analysis of variance shown in Table 2 indicate that the three diagnostic groups were significantly

TABLE 1

WEIGHT DISCRIMINATION OF SCHIZOPHRENIC AND NORMAL GROUPS

Group	Mean percentage of correct discriminations		Mean weber scores $\left(\frac{\Delta S}{S}\right)$	
	Light[a]	Heavy[b]	Light[a]	Heavy[b]
Disturbed chronic schizophrenics				
Heavy-Light ($N = 6$)	65.4	75.0	.283	.142
Light-Heavy ($N = 6$)	64.2	80.0	.292	.158
All subjects ($N = 12$)	64.8	77.5	.287	.150
Chronic schizophrenics				
Heavy-Light ($N = 10$)	72.0	82.8	.160	.107
Light-Heavy ($N = 10$)	76.3	91.5	.137	.062
All subjects ($N = 20$)	74.1	87.1	.149	.085
Normals				
Heavy-Light ($N = 10$)	81.8	88.5	.120	.070
Light-Heavy ($N = 10$)	85.8	94.0	.087	.057
All subjects ($N = 20$)	83.8	91.2	.104	.064

[a] 40-Gm. Std.
[b] 400-Gm. Std.

different in their mean percentage of correct discriminations ($p < .001$) for the combined weight intensities. Separate t tests showed that normals performed significantly better than both groups of schizophrenics and that the chronic group was superior to the chronic disturbed group (all p values $< .01$). It is also apparent that all groups discriminated better on the heavy series of weights ($p < .001$). While there was some tendency for all subjects to perform better when tested in the light-heavy order, this effect was not statistically significant ($p < .10$). The interaction of order and weights was significant ($p < .01$), however, indicating that the preferred order improved heavy discriminations more than light weight discriminations. The heavy weights appear to have produced a more durable anchor effect which interfered with practice effects in the heavy-light order, while benefiting from practice in the light-heavy order.

TABLE 2

ANALYSIS OF VARIANCE OF WEIGHT DISCRIMINATION SCORES

Source	df	Percentage of correct discrimination		Weber scores	
		MS	F	MS	F
Diagnosis (D)	2	2.68	16.75**	7.80	21.08**
Order (O)	1	.50	3.13	.65	1.76
D × O	2	.18	1.13	.26	—
Error (b)	46	.16		.37	
Total Between	51				
Weights (W)	1	3.40	87.18**	8.71	28.46**
W × D	2	.27	6.92*	.81	2.65
W × O	1	.40	10.26*	.08	—
W × D × O	2	.06	1.41	.32	1.06
Error (w)	46	.04		.31	
Total	103				

* $p < .01$.
** $p < .001$

The main hypothesis of this study was that schizophrenics would approach normal discrimination with the heavier weights. A significant interaction between weights and diagnosis ($p < .01$) is in keeping with this hypothesis. Inspection of the table shows that both groups of schizophrenics improve approximately 13% on the heavier weights, while normals show only 7.4% improvement. In addition, the chronic group closely approximates normal discrimination on the heavy series and is not statistically differentiable.

Weber ratio scores. Examination of the mean Weber scores in Table 1 and the results of the analysis of variance show essentially the same results obtained for percentage of correct discriminations. The main effects for diagnosis and weights were again highly significant (both p values $< .001$), while the order effects were not statistically significant on this measure. The interaction between diagnosis and weight intensities was not quite significant ($p < .10$), but the same

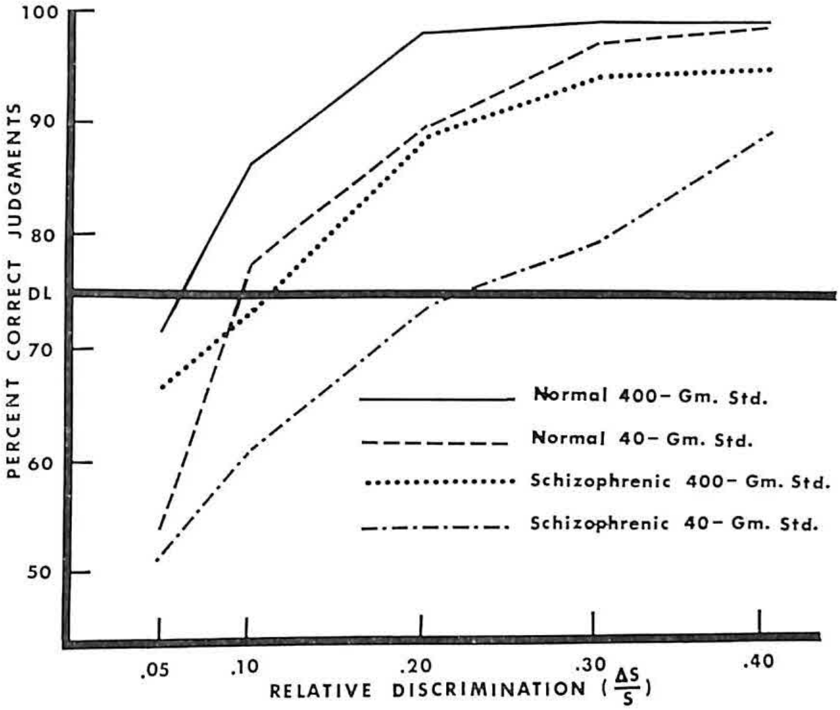


FIG. 1. Weight discrimination of normals and schizophrenics at two intensities.

trends are apparent in the table. Both schizophrenic groups show large improvements in Weber rations on the heavy series of weights and the chronic group is not significantly different from normals at the heavy intensity.

The effects of weight intensity on normal and schizophrenic discrimination thresholds for all patients combined is shown graphically in Figure 1. Examination of the curves shows that ΔS/S for normal subjects was approximately .10 or 4 grams at the difference limen (DL) for the 40-grams standard. For the combined schizophrenics ΔS/S approaches .23 or 9.2 grams for the 40-grams standard at the DL. It is interesting to note that both these values correspond very closely to those obtained from normals and schizophrenics in the Sernyl study cited earlier (Rosenbaum et al., 1959). Inspection of the curves for the 400-grams standard shows that the schizophrenic curve closely resembles that for normals and crosses the DL when ΔS/S approaches .11, approximating normal discrimination thresholds.

DISCUSSION

The present results confirm previous findings with chronic schizophrenics showing marked deficits in discrimination thresholds for lifted weights. The extent of this deficit was found to be related to severity of pathology in that the chronic disturbed group performed significantly more poorly than the less disturbed chronic patients. In keeping with the major hypothesis of the present study, both groups of schizophrenics showed marked improvements in discrimination thresholds with the heavy weights.

Salzinger's (1957) failure to obtain a difference between acute schizophrenics and normals in their absolute judgments of 200-to-400-grams weights appears to be explicable in terms of the present results. Since the less disturbed chronic schizophrenic group in this study did not differ significantly from normals in the 400-grams series, it seems unlikely that an even better integrated acute group of patients would show impairment on the heavy series of weights that he employed. In addition, the mean Weber ratio of .085 obtained from our chronic schizophrenics on the heavy series (showing a discrimination threshold of approximately 35 grams) suggests that absolute judgments of weights differing by 50 grams in this range of intensities might well be above the discrimination thresholds of acute schizophrenics. While it is also possible that acute schizoassociative behavior scores in terms of (*a*) "over-all" associative behavior, (*b*) "most common" associative behavior, and (*c*) "least common" associative behavior. The greater the discrepancy between the subject's performance and that of the norm sample, the greater was the associative disturbance attributed to the subject.

RESULTS AND DISCUSSION

The maximum possible "over-all" associative behavior score was 10,424. The subjects achieved a mean "over-all" score of 4,820 ($SD = 2{,}085$, range = 308–7,747). The relationship between "over-all" score and length of hospitalization was in the predicted direction and statistically significant ($r = -.42$, $df = 45$, $p < .01$).

The subjects gave the "most common" response to the 50 stimulus words a mean of 18 times ($SD = 8$, range = 1–30). The relationship between "most common" score and duration of hospitalization was also in the predicted direction and significant ($r = -.40$, $df = 45$, $p < .01$).

With respect to idiosyncratic responses to the 50 stimulus words, the "least common" associative behavior mean score for the subjects was 10 ($SD = 7$, range = 2–35). The relationship between "least common" score and hospitalization was again in the predicted direction and significant ($r = +.45$, $df = 45$, $p < .01$).

When age was held constant through partial correlation, the above coefficients became $-.41$, $-.39$, and $+.45$, respectively.

All three word association measures, that is, "over-all" associative behavior, "most common" associative behavior, and "least common" associative behavior provided support for the hypothesis of greater associative disturbance with increasing chronicity in schizophrenia. The results were thus consonant with both clinical observation and Mednick's analysis of the course of the disorder. It was therefore concluded that the results suggest that associative disturbance is more prevalent in the chronic stages of schizophrenia than in the acute stages.

REFERENCES

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: Int. Univer. Press, 1950.

MEDNICK, S. A. A learning theory approach to research in schizophrenia. *Psychological Bulletin,* 1958, **55,** 316–327.

MEDNICK, S. A. Learning theory and schizophrenia: A reply to a comment. *Psychological Bulletin,* 1959, **56,** 315–316.

MEDNICK, S. A. Schizophrenia: A learned thought disorder. In G. S. Nielsen (Ed.), *Proceedings of the XIV International Congress of Applied Psychology.* Vol. 4. *Clinical psychology.* Copenhagen: Munksgaard, 1962. Pp. 167–178.

PALERMO, D. S., & JENKINS, J. J. *Word association norms: Grade school through college.* Minneapolis: Univer. Minnesota Press, 1964.

(Received February 5, 1965)

---

through college; this condition not holding, responses which were not given by any member of the college-male group.

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 453–456

SOME EFFECTS OF LSD-25 ON VERBAL COMMUNICATION

MARIANNE AMAREL AND FRANCES E. CHEEK

New Jersey Bureau of Research in Neurology and Psychiatry, Princeton

The effect of LSD-25 on 2 dimensions of verbal behavior, predictability and volubility, was examined. 100-word samples of spontaneous speech from 10 alcoholic Ss under 4 drug conditions (predrug, 100 μg., 200 μg., postdrug) were elicited. 100 normal raters were asked to fill in every 5th-deleted word of the speech samples, all raters completing a set of 4 samples from randomly ordered Ss. The communicative value of speech (ratio of total to correct guesses) was found to decrease under LSD, as did volubility (number of words spoken). Additionally, an inverse relationship between predictability and volubility under the drug was suggested.

Investigators have reported remarkable alterations in personality, often judged therapeutically beneficial, resulting from the administration of the psychotogenic drug LSD-25 (Unger, 1963). These changes have alternatively been attributed in part to enhanced transference (Schmiege, 1963), heightened degree of empathy (Osmond, 1957), and similar factors relating to the communication process.

A study of the effect of LSD on the communication process in small group settings was undertaken by Lennard, Jarvik, and Abramson (1956). Using formal characteristics of communication such as quantity and direction of speech, and also the Bales Interaction Process Analysis categories, the authors found that nonpsychiatric subjects under LSD produce less speech, and that of lower density, with a greater number of unfinished sentences and interruptions. In a later study, Abramson, Hewitt, Lennard, Turner, O'Neill, and Merlis (1958) found that a schizophrenic patient in a three-person group including the schizophrenic, a stablemate (i.e., a specially selected normal with whom the patient could identify), and an interviewer, increased his communication output relative to the stablemate when he was administered LSD. Another investigator (Cheek, 1963), observing a four-man group, found the total interactions of the group increased under LSD, at dosages of 25 to and including 100 micrograms.

In these studies, communication was studied as a group process. In the work to be reported here, the informational value of speech, rather than the pattern of communicative acts, was examined. The communicator's, or source's, output was evaluated in terms of its effect on the communicand, or the target. To measure the communicative value of speech, the cloze procedure (Taylor, 1953) was used. This technique involves the deletion of every *n*th word in a verbal passage, after which raters are asked to fill in the missing word with one that seems best to fit the remaining context. The number of correctly guessed words thus forms a measure of the predictability of the given verbal sample.

The cloze procedure has been used in a variety of contexts. It has been found to differentiate the speech of aphasics from that of normal subjects (Fillenbaum & Jones, 1962), and schizophrenic from normal speech (Salzinger, Portnoy, & Feldman, 1964) It has also been used to gauge the effect of psychotherapeutic interaction on predictability (Feldstein & Jaffe, 1963).

In the area directly relevant to the present effort, the cloze procedure has been used to measure the effect of drugs on predictability of speech. Salzinger, Pisoni, Feldman, and Bacon (1961) reported that the speech of one subject became less predictable with increasing dosages of chlorpromazine, while Honigfeld (1963), using psilocybin, again with one subject, found over a 3-hour period after drug ingestion, predictability first enhanced, later decreased.

The authors of the present study conducted a pilot investigation, examining the speech of one subject in a four-person discussion group, all subjects being under various drug conditions (placebo, 25 μg., 75 μg. and 100 μg. of LSD-25), finding predictability decreasing with mounting drug dosage.

However, in all three studies the speech of only one subject was examined, with experimental conditions sufficiently different to make comparisons across studies tenuous. In the present study the sample size of subjects as well as raters was increased, and the conditions under which the speech was elicited specified to facilitate replication.

METHOD

The subjects were 10 male patients from the Alcoholic Treatment Center of the New Jersey Neuro-Psychiatric Institute participating in a program in which LSD-25 was used as a psychoadjuvant to group therapy. The subjects were diagnosed as alcoholics with no known organic involvements, ranged from 29 to 63 years in age, with a median age of 37 years. Their education ranged from 8th grade to 4 years of college, with a median of 12th grade.

The drug was administered on two separate occasions, a low dose (100 μg.) followed by a high dose (200 μg.) 1 week later, at 8:30 A.M. Spontaneous speech of the subjects was recorded at 1:00 P.M. under both of these drug conditions, as well as during a predrug session 1 week before the first drug administration, and a postdrug session 1 week after the second drug administration.

The speech was elicited in the following manner. The subject was comfortably seated, facing a mirror, with a tape recorder in view. On each occasion the interviewer used these instructions:

> I am going to ask you to tell me something about yourself. I would like you to look into this mirror and for 5 minutes say aloud whatever thoughts about yourself come to your mind. I don't care what you talk about—your appearance, what sort of person you are, how you get along with other people, how other people might think or feel about you, just so long as you talk about yourself. Are there any questions?

The interviewer then left the room, returning in 5 minutes, saying:

> Now I am going to ask you to tell me something about your family. For 5 minutes I would like you to say aloud whatever thoughts about them that come to your mind. You can talk about whatever you want, what sorts of people they are, how you all get along, how you might think or feel about them, just so long as you talk about your family. By family, I mean either your parents or your wife and children, or both. Are there any questions?

Leaving the subject again, the interviewer returned once more after 5 minutes, this time asking the subject:

> Now I am going to ask you to tell me something about this group with whom you are taking LSD. For 5 minutes I would like you to say aloud whatever thoughts come to your mind about this group. You can talk about whatever you want, what sorts of people they are, how you all get along with one another, how you might think or feel about them, just so long as you talk about the group. Are there any questions?

Thus the subjects produced a total of three 5-minute segments of speech during each session.

The taped speech was transcribed, including all repetitions, uncompleted words, sighs, etc. The first 100 words of the first segment (self-description) of each of the 4 monologues from all 10 subjects were selected, representing the 4 experimental conditions. From these speech samples broken words and other speech disruption were deleted, but repetitions were retained.

The samples were then prepared by the deletion of every fifth word, with a space of uniform length substituted. Wherever the fifth word was a proper noun or number, the next word was deleted with a corresponding number of words added to the passage, so that the last twentieth blank was always followed by four words.

The four speech samples of 10 subjects were duplicated 10-fold to allow for the construction of 100 booklets containing 4 speech samples each. Each rater thus filled in one sample of each drug condition randomized with regard both to subjects and order of drug condition.

The raters used were 100 female members of the Junior class of Trenton State College (age range 19–21) who received the following instructions in a group session:

> This is a study of language. On each of the next four pages of your booklet is a sample of a person's speech. All punctuation has been omitted. You will notice that every so often there is a blank; every fifth word which was spoken has been left out. Your job is to fill the blank with the word you think will make the most sense. The missing words are *not* numbers or names. This is difficult and often you will have to guess. Do the best you can. It is *very important* that you fill in *all* the blanks. Work as quickly as you can.

Most raters completed the task in less than the allotted 30 minutes. The data consisted of the number of guessed words identical to those recorded.

RESULTS AND DISCUSSION

Predictability and Drug Condition

Table 1 presents the mean cloze scores (ratio of correct to total guesses) of the 10 ratings obtained for each speech passage of the 10 alcoholic subjects in each of the 4 drug conditions.

The predictability of speech appears to decline when the subject is under LSD-25. The mean cloze scores show greatest predictability in the predrug phase, then the postdrug phase. The high dose condition is least predictable, and the low dose condition only slightly more so.

Table 2 shows the results of an analysis of variance (performed on the original measures) investigating the influence on predictability of the drug conditions and the subject giving the spontaneous speech sample, and the interaction of these two factors.

Evidently, the drug condition has a significant influence upon the predictability of the speech samples, for the speech of the 10 subjects is significantly ($p < .01$) more difficult to predict when they are under the effect of LSD-25.

Also, the speech of some subjects is more difficult to predict than the speech of others ($p < .01$). The interaction F ratio is also significant ($p < .01$), indicating that the predictability of some subjects is more affected by the drug condition than the predictability of others.

TABLE 1

Mean Cloze Scores of Ten Ratings of Speech Samples of Ten Subjects under Four Drug Conditions

Subject	Drug condition			
	Predrug	Low dose (100 μg)	High dose (200 μg)	Postdrug
A	.40	.26	.27	.47
B	.46	.38	.47	.33
C	.61	.30	.33	.37
D	.61	.29	.34	.32
E	.48	.44	.65	.54
F	.42	.50	.39	.32
G	.31	.47	.26	.43
H	.45	.39	.29	.34
I	.37	.43	.25	.63
J	.31	.27	.33	.36
Grand M	.44	.38	.36	.41

TABLE 2

Analysis of Variance of Differences in Cloze Scores in Relation to Drug Conditions and Subjects

Source	df	MS	F
Subjects (A)	9	49.69	12.43*
Drug conditions (B)	3	50.96	12.68*
A × B	27	31.37	9.84*
Within treatments	360	4.00	
Total	399		

* $p < .01$.

TABLE 3

Verbal Output (Number of Words) of Ten Subjects under Four Drug Conditions

Subject	Drug condition			
	Predrug	Low dose (100 μg)	High dose (200 μg)	Postdrug
A	1,664	909	1,230	1,894
B	1,220	1,181	291	1,259
C	1,069	1,224	1,017	1,477
D	2,129	1,569	1,005	1,762
E	985	270	264	836
F	1,670	337	540	1,806
G	1,220	553	485	967
H	2,127	1,692	1,568	2,277
I	1,869	1,659	1,243	1,671
J	2,502	2,154	1,983	2,462
M	1,645.5	1,154.8	962.6	1,641.1

TABLE 4

Anatysis of Variance of Differences in Verbal Output of Ten Subjects in Four Drug Conditions

Source	df	MS	F
Between conditions	3	1,303,392	4.67*
Within sets	36	278,808	
Total	39		

* $p < .01$.

Volubility and Drug Conditions

Another dimension of the subjects' verbal behavior was examined in relation to the drug conditions, that of volubility, defined as quantity of verbal output per unit of time. A measure of the total verbal output for each subject under each condition was established by summing the words of the typescripts yielded by the three 5-minute monologues.

Table 3 presents the resulting figures, showing verbal output to be highest in the predrug conditions, decreasing with increasing drug dosage, returning to near predrug level in the postdrug session. An analysis of the variance due to the difference in drug conditions, presented in Table 4, shows the drug conditions to have a significant ($p < .01$) effect on verbal output. All subjects speak less while under the influence of the drug.

TABLE 5

CORRELATIONS BETWEEN CLOZE SCORES OF RATERS AND VERBAL OUTPUT OF SUBJECTS IN FOUR DRUG CONDITIONS

Drug condition	ρ[a]
Predrug	−.24
Low dose—100 μg	−.53
High dose—200 μg	−.54
Postdrug	−.30

[a] .60 significant at the 5% level.

Predictability, Volubility, and Drug Conditions

Next, the relation between predictability and volubility was examined. Table 5 shows the rank order correlations between amount of verbal output and cloze scores of passages for each subject in the four drug conditions.

All four coefficients are negative; two approach significance at the $p = .05$ level, suggesting an inverse relationship between volubility and predictability when the drug effect is present. It appears that those subjects who are more voluble under the drug are also less predictable under the drug. However, this relation between volubility and predictability is weaker when the subjects are free of drug effect. It may be that talkativeness under the drug is associated with loss of control and a loosening of associative bonds. Further examination of the speech protocols in order to investigate this finding is now in progress.

Stability of Volubility and Predictability

Stability of the subjects through the various conditions in terms of volubility and predictability was also examined. The verbal output scores of each subject in the two nondrug and drug conditions were summed, rank ordered, and correlated, yielding a rho of .85, which is significant above the .01 level, indicating that subjects maintain their relative position of volubility to a high degree across conditions. The subjects' cloze scores were similarly treated, yielding a rho of .06, which is not significant, suggesting that predictability is a characteristic of speech more variable and sensitive than is volubility. This assumption is somewhat strengthened by the fact that all subjects lowered their verbal output under LSD-25, whereas three subjects deviated from the general trend by rising in predictability under the drug. The cloze technique may be tapping multiple factors differentially affected by the drug conditions.

Thus, our findings support those of our own pilot studies and those of other investigators that communication as measured by predictability of speech and amount of verbal output is reduced by the administration of LSD-25. In view of this, it is interesting that the therapeutic effect of the drug has been attributed to facilitated communication and increased empathy.

REFERENCES

ABRAMSON, H. A., HEWITT, M. P., LENNARD, H., TURNER, WO. S., O'NEILL, F. J., & MERLIS, S. The stablemate concept of therapy as affected by LSD in schizophrenia. *Journal of Psychology*, 1958, **45**, 75–84.

CHEEK, F. E. Exploratory study of drugs and social interaction. *Archives of General Psychiatry*, 1963, **9**, 566–574.

FELDSTEIN, S., & JAFFE, J. Language predictability as a function of psychotherapeutic interaction. *Journal of Consulting Psychology*, 1963, **27**, 123–126.

FILLENBAUM, S., & JONES, L. V. An application of "Cloze" technique to the study of aphasic speech. *Journal of Abnormal and Social Psychology*, 1962, **65**, 183–189.

HONIGFELD, G. Effect of an hallucinogenic agent on verbal behavior. *Psychological Reports*, 1963, **13**, 383–385.

LENNARD, H., JARVIK, M. E., & ABRAMSON, H. A. Lysergic acid diethylamide (LSD-25): XII. A preliminary statement of its effects upon interpersonal communication. *Journal of Psychology*, 1956, **41**, 185–198.

OSMOND, H. A review of the clinical effects of psychotomimetic agents. *Annals N. Y. Academy of Science*, 1957, **66**, 418–434.

SALZINGER, K., PISONI, S., FELDMAN, R. S., & BACON, P. M. The effect of drugs on verbal behavior. In Control of verbal behavior. Symposium presented at the American Association for the Advancement of Science, Denver, December 1961.

SALZINGER, K., PORTNOY, S., & FELDMAN, R. S. Verbal behavior of schizophrenic and normal subjects. *Annals of the N. Y. Academy of Sciences*, 1964, **105**, 845–860.

SCHMIEGE, G. R. The current status of LSD as a therapeutic tool—A summary of the clinical literature. *Journal of the Medical Society of New Jersey*, 1963, **60**, 205–207.

TAYLOR, W. L. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 1953, **30**, 415–433.

UNGER, S. M. Mescaline, LSD, psilocybin, and personality change. *Psychiatry, Journal for the Study of Interpersonal Processes*, 1963, **26**(2), 111–125.

(Received February 17, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 457–461

SOCIAL SCHEMATA OF EMOTIONALLY DISTURBED BOYS[1]

LAURA WEINSTEIN

George Peabody College for Teachers

2 experiments, using Kuethe's techniques, were designed to demonstrate that children defined as emotionally disturbed differ predictably from normal children in the ways they organize social stimuli. When emotionally disturbed and normal boys placed pairs of felt figures on flannel boards, normals placed child figures closer to mother than to father or peer figures; emotionally disturbed children did the reverse. When disturbed and normal boys replaced pairs of human and geometric figures previously set 15 ins. apart, the disturbed boys replaced the human figures farther apart than the nonhuman figures significantly more often than the normals. Results were interpreted as reflecting a tendency for emotionally disturbed children to construe people, and especially females, more negatively than do normal children.

There is considerable consensus in contemporary psychology about the importance of cognitive mediation in behavior, the way the individual behaves being seen as a function of the way he organizes and interprets the world about him. Following the logic of such approaches, it should be possible to relate deviant behavior to deviations in these mediating processes. As a step in this direction, the present study compares the ways emotionally disturbed and normal children organize socially meaningful stimuli.

Kuethe and his associates (Kuethe 1962a, 1962b, 1964; Kuethe & Stricker, 1963; Kuethe & Weingartner, 1964) have shown that when normal adult subjects place human figures on a field there is great similarity in the organizations they produce. As examples, human figures are kept together, unseparated by nonhuman objects; male figures are placed near female figures with same-sex figures rarely being paired; the figure of a child is placed closer to a mother than a father figure. Similarly, when the subjects are told to replace figures after viewing their positions for a short time, they err in directions consistent with the commonly produced free placements. Patterns of verbal association are also correlated with placement of figures.

Kuethe accounts for these consistencies by positing the existence of common social schemata. Commonalities in the past experience of a cultural group produce shared schemata which provide individuals with common frames of reference for organizing and responding to social stimuli.

Kuethe goes on to examine whether deviant groups have social schemata which differ from those of normals and which are related to the nature of their deviance. He finds this to be the case for homosexuals who, in replacing male pairs closer than mixed pairs, utilize a schema consistent with their sexual orientation and different from that of nonhomosexuals.

This study is an extension of Kuethe's analysis using a different deviant group, emotionally disturbed children. Their deviant behavior should also be associated with relevant deviant social schemata.

Two aspects of their social schemata are of particular interest. First, like normal adults, do emotionally disturbed children group humans closer than nonhuman objects? Or, have they learned that humans are punishing, and therefore place them farther than neutral stimuli? Second, a disturbed mother-child relationship is commonly regarded as having etiological significance in children's

---

[1] This investigation was supported by Public Health Service Grant No. MH 929 from the National Institute of Mental Health. Subjects were obtained from Cumberland House and Wright School, two residential elementary schools for the re-education of emotionally disturbed children sponsored by Peabody College and the states of Tennessee and North Carolina. The author wishes to thank the staffs of these schools and also Hoyte M. Snow, public school principal, and his teaching staff for their cooperation and help. Special thanks are due Evelyn Cromwell, Barbara Gilmer, Miranda Godwin, and Judith Snellgrove for their help in collecting the data.

TABLE 1

AGE, IQ AND GRADE LEVEL OF GROUPS IN EXPERIMENTS I AND II

Group	Experiment I							Experiment II						
	N	Age in months		IQ[a]		Grade level		*N*	Age in months		IQ[b]		Grade level	
		M	*SD*	*M*	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*	*M*	*SD*
Emotionally disturbed	20	123.6	12.8	94.4	17.3	4.0	1.1	25	126.1	14.9	101.5	14.0	3.9	1.5
Normal	20	123.8	12.2	95.5	15.8	4.0	1.0	39	119.9	15.9	99.7	12.7	4.3	1.0

Note.—None of the differences between groups were significant by t test.
[a] WISC for the disturbed children, Lorge-Thorndike for normals.
[b] WISC or PPVT for the disturbed children, Lorge-Thorndike for normals.

emotional difficulty. Is this disturbance reflected in the figure placements of the emotionally disturbed child?

EXPERIMENT I

Method

Subjects. The subjects were 40 boys ranging in age from 8 to 12 years. Twenty were consecutive white male admissions to the main campus of a residential school for emotionally disturbed children. The experiment was limited to white males because the number of Negroes and girls admitted to the school during this period was too small to make tests of sex or race differences meaningful. For a similar reason, seven extremely socially deviant boys admitted to a special camp program were omitted from the experiment. The camp group contained a number of boys who might be classified as psychopaths, while the campus children could generally be classified under the broad rubric "neurotic." Since the different labels may reflect differences in social schemata, the experimental sample was limited to the campus children, who appeared more representative of the general meaning of "emotionally disturbed." The symptoms of this latter group varied from acting out to withdrawal, often included psychosomatic symptoms and almost always included learning problems. Generally these boys were tested within 2 weeks after admission. The 20 control children were boys from a local public school matched with the disturbed group on age, IQ, and grade level. Table 1 indicates that matching was close on all three dimensions.

Procedure. A variant of Kuethe's felt figure free placement technique was used. The subjects were asked to place five pairs of figures on individual flannel boards. When placement of one pair of figures was completed, the flannel board was removed and another blank board was presented along with a separate envelope containing the next pair of figures. This permitted measuring to take place after the subject had left the experimental room. The boards and figures were somewhat smaller than Kuethe's. The boards were 14½ inches high and 24 inches wide. The large figures were 7 × 2¾ inches and the small figures 5¼ × 2½ inches.

The five pairs of figures were as follows: a mother figure and a child figure, a father figure and a child figure, two child figures, two rectangles the size of the child figures, one rectangle equal in size to a parent figure and one equal to a child figure. Order of presentation was randomized for both groups. Comparisons were made of the relative distances between selected pairs.

Results

Mother-Child Relationship. Table 2 compares the relative placements of the mother-child and father-child pairs by the two groups of children. Kuethe finds that the majority of normal adults place child figures closer to adult female than adult male figures. So did this sample of normal children. The disturbed children, in significant contrast (χ^2 corrected for continuity $= 3.608$, $p < .05$ one-tailed test), placed the child closer to the adult male.

Does this difference reflect a particular closeness of father and child in the social schemata of emotionally disturbed children, or are the results due rather to rejection of the mother-child relationship? The results in Table 2 comparing the mother-child and father-child pairings with the child-child pairings shed light on this point. These data go beyond Kuethe's by considering a third alternative, the peer relationship. This relationship can be used as a benchmark in assessing the mother-father comparison.

Table 2 shows both the emotionally disturbed and normal groups about evenly balanced in placing either fathers or peers closer. Thus, any interpretation of the mother-father comparison as due to the inordinate attraction of the father for disturbed children or

TABLE 2

RELATIVE PLACEMENTS OF FIGURE PAIRS

Group	Experiment I: Free placement								Experiment II: Replacement	
	Mother-Child versus Father-Child[a]		Father-Child versus Child-Child[b]		Mother-Child versus Child-Child[c]		Humans versus Rectangles		Humans versus Rectangles	
	MC Closer	FC Closer	FC Closer	CC Closer	MC Closer	CC Closer	HH Closer	RR Closer	HH Closer	RR Closer
Emotionally disturbed	6	14	12	8	4	16	3	17	5	20
Normal	13	7	9	11	12	8	6	14	19	20
	$\chi^2 = 3.608$, $df = 1$, $p < .05$, one-tailed		$\chi^2 = .401$, $df = 1$, *ns*		$\chi^2 = 5.102$, $df = 1$, $p < .05$		$\chi^2 = .573$, $df = 1$, *ns*		$\chi^2 = 4.271$, $df = 1$, $p < .05$	

Note.—The chi-square tests incorporated a correction for continuity.

[a] Two children in the disturbed group placed the child figure equally close to the two parent figures. Each tie was resolved by assigning a frequency of .5 to each category for the disturbed group. Omitting the ties accentuates the contrast between the groups.

[b] There were four ties in the disturbed group. Ties were resolved as indicated above.

[c] There were two ties in the disturbed group. Ties were resolved as indicated above.

rejection of him by normals appears unlikely. Nor do peers appear to have a unique attraction for either group. The significant χ^2 (5.102 corrected for continuity, $p < .05$) for the mother-child versus child-child comparison, on the other hand, supports the interpretation of rejection of the mother in the social schemata of disturbed children. Mothers are placed closer than peers by 60% of the normal children, but by only 20% of the disturbed children.

Human-Nonhuman Relationship. Table 2 compares the relative distances of the three human with the two nonhuman placements, using the mean of each type for each subject. Contrary to expectation, both groups placed the rectangles closer than the human figures. Results for this comparison were contaminated, however, by the fact that a number of children in each group "made something" out of the rectangles—usually a "T" or a cross. Because this made it difficult to interpret the human-nonhuman results, a second test of this relationship was undertaken.

EXPERIMENT II

The second experiment utilized Kuethe's replacement technique. This technique directs the subject to reproduce accurately an organization created by the experimenter, rather than to create one himself. The emphasis on accuracy rather than creativity effectively eliminates the contamination mentioned above.

Method

Subjects. The subjects for the disturbed group in the second experiment were the 25 white boys enrolled in the two residential schools for emotionally disturbed children on the day the experiment took place. The children were therefore in various stages of treatment. Eleven of the 25 emotionally disturbed subjects used in this second experiment had previously participated in the first. There was no other overlap of subjects in the two experiments. The 39 subjects in the normal group consisted of third and fifth grade boys from the public school used in the first experiment who were present on the day of the experiment and for whom IQ scores were available. As a partial check on the "normality" of this control group, teachers were interviewed about the possibility of emotional disturbance among the children to be tested. In only two cases was the evidence sufficiently clear-cut to eliminate children from the sample. However, there were a number of doubtful cases, as one might expect in a school such as the one sampled, which draws heavily from lower socioeconomic status families. The decision to retain the remainder of the sample as the control group was based on the fact that many of its less desirable family, IQ, academic achievement, and behavioral characteristics are also largely true of the disturbed group, although not necessarily associated with emotional disturbance.

Subjects were between the ages of 7 and 12. Age, IQ, and grade level statistics are reported in Table 1.

Procedure. When the subject entered the room, he was asked to look for 5 seconds at either a pair of rectangles or a pair of human figures (a male and a female) placed 15 inches apart on a large piece of felt fastened to the wall 10 feet in front of him. The figures were of white felt and were approximately 10½ inches high and 3½ inches wide; the background felt was black, 2 yards wide and 1½ yards high. The subject was told that when the figures were taken down and given to him, he was to replace them exactly as far apart as they had been. The subject left the room while the experimenter measured the replacement and put up the second pair of figures. All subjects replaced both pairs, with order of presentation counterbalanced over subjects. Measurements were made from midline to midline. The experiment was presented to the children as a study of accuracy in judging distances.

Results

Table 2 compares the performance of the two groups. The great majority of the emotionally disturbed group (20 out of 25, $p < .002$, one-tailed test) replaced the human figures farther apart than they replaced the rectangles. The normal group made no systematic distinction between the two kinds of figures. That the two groups responded differently to the two kinds of figures is indicated by the significant χ^2 of 4.271, $p < .05$.

It had been predicted that the disturbed group would react in a manner opposite to Kuethe's sample of normal college students, and this prediction was confirmed. The normal children, however, although they reacted differently from the disturbed children, did not make the human-nonhuman distinction made by Kuethe's adult sample. Two possible explanations come to mind. First, a developmental factor may be operating so that clear-cut schemata do not normally emerge until after the ages dealt with in this study (perhaps a latency effect). Second, the normal sample, as previously noted, may not have been as homogeneously normal as Kuethe's college adults. Further research is needed to clarify this point.

DISCUSSION

Kuethe's free placement technique, like the projective devices, assumes that unstructured situations elicit schemata (response sets, meanings, personal constructs) which typically operate. The technique is particularly valuable in studying the degree to which objects are construed as belonging together. With it, Kuethe has identified a number of schemata which reflect cultural norms and are widely shared among normal adults. For example, human beings, and particularly male-female and mother-child dyads, are viewed as highly related by most adult subjects. The replacement studies have shown that the same schemata lead to predictable errors of judgment. They appear also to determine word associations; subjects who keep mixed sex pairs together in free placement also give "man" and "woman" as reciprocal verbal associations. His work with the homosexual sample indicates that these schemata also mediate socially important behaviors.

Kuethe's adult samples did not permit a nonhuman object to intervene between human figures; the same schema caused them to replace human figures too close together. These findings are consistent with results based on subjective probabilities in minimal information situations (DeSoto & Kuethe, 1959). Given only the information that the people are acquainted, subjects think it probable that the people like each other. Human beings are generally seen as related, friendly, prone to interact.

Our results indicate that emotionally disturbed children have not developed the normal adult schema which organizes humans as a close unit. Whereas adults replace humans closer than rectangles, emotionally disturbed children replace the rectangles closer. It is uncertain whether the disturbed children have developed an opposite schema, one which conceives humans as separate and unfriendly, or have simply not developed a positive human relationship schema, since closer placement of the rectangles might be expected on the basis of the gestalt unit-forming principle "similarity" (Koffka, 1935). The rectangles were identical in size and shape, while the male and female figures differed in shape. If the similarity principle operated for the adults, it was less powerful than the schema that people belong together. In the sample of normal children, the two forces may have canceled each other out.

The research on parent-child attitudes gives indirect support to the hypothesis that emotionally disturbed children construe human beings as more separate and more negative.

For example, Cox (1962) reports that boys described by their peers as aggressive reject one or both of their parents (rejection was determined from TAT stories) while Schulman, Shoemaker, and Moelis (1962) find parents of children with conduct problems more hostile and rejecting in observed parent-child interactions than parents of normal children. This construction of human beings and human relationships might be expected to vitiate the effectiveness of social approval, upon which socialization so largely depends, as a reinforcer for these children. Levin and Simmons (1962) have reported that in a free operant, two-choice marble-dropping task, praise did not function as a reinforcer for severely emotionally disturbed children. It would be interesting to know whether pattern of operant response under Levin and Simmon's experimental conditions would be related to the replacement behavior described here.

Our emotionally disturbed sample, in contrast to the normal children, placed mother-child pairs farther apart than either father-child or child-child pairs. The behavior of the disturbed children is also unlike that of Kuethe's adults, who placed mother-child pairs closer than father-child pairs and who, in forming father-mother-child triads, rarely put the father between the mother and child. Neither is it consistent with the findings of Kagan and Lemkin (1960), who report that children see mothers as nicer and more apt to give presents, fathers as more punitive and fear arousing. The schemata of emotionally disturbed children appear to single out the woman-child relationship as particularly distant and negative. This may be because these relationships are primary trouble spots; it may reflect greater opportunity for negative interaction because of the relatively large amount of time a child spends with mother and (usually female) teacher. At any rate, it probably reflects the great saliency of mothers for young children.

If we can equate physical distance placed between human figures with emotional distance (rejection, withdrawal, distrust, and their opposites) as is often done colloquially ("we are close friends"; "she was very distant with me"; "keep him at arm's length") and clinically, in interpreting doll play, drawings, and other projective devices, emotionally disturbed children construe humans, and particularly adult females, more negatively—perhaps as more punitive or less supportive and less trustworthy. Two important questions are raised for further research. First, do these differences have implications for the effectiveness of social approval as a reinforcer for these children, particularly approval from the primary sources of behavior shaping in young children, mothers, and teachers? Second, is effective treatment for these children associated with a change in social schemata as reflected in the techniques used in these experiments?

REFERENCES

COX, F. N. An assessment of children's attitudes towards parent figures. *Child Development,* 1962, 33, 821–830.

DESOTO, C., & KUETHE, J. L. Subjective probabilities of interpersonal relationships. *Journal of Abnormal and Social Psychology,* 1959, **59**, 290–294.

KAGAN, J., & LEMKIN, JUDITH. The child's differential perception of parental attributes. *Journal of Abnormal and Social Psychology,* 1960, **61**, 440–447.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

KUETHE, J. L. Social schemas. *Journal of Abnormal and Social Psychology,* 1962, **64**, 31–38. (a)

KUETHE, J. L. Social schemas and the reconstruction of social object displays from memory. *Journal of Abnormal and Social Psychology,* 1962, **65**, 71–74. (b)

KUETHE, J. L. The pervasive influence of social schemata. *Journal of Abnormal and Social Psychology,* 1964, **68**, 248–254.

KUETHE, J. L., & STRICKER, G. Man and woman: Social schemata of males and females. *Psychological Reports,* 1963, **13**, 655–661.

KUETHE, J. L., & WEINGARTNER, N. Male-female schemata of homosexual and non-homosexual penitentiary inmates. *Journal of Personality,* 1964, **32**, 23–31.

LEVIN, G. R., & SIMMONS, J. B. Response to praise by emotionally disturbed boys. *Psychological Reports,* 1962, **11**(1), 10.

SCHULMAN, R. E., SHOEMAKER, D. J., & MOELIS, I. Laboratory measurement of parental behavior. *Journal of Consulting Psychology,* 1962, **26**, 109–114.

(Received March 4, 1965)

scoring in the high range on Welsh's *A* Scale will give significantly more doubtful judgments than subjects scoring in the low range on the same scale, was therefore fully confirmed.

Differences in Sensitivity between Groups

In order to establish anxiety level as the crucial variable in the differential use of the judgment category doubtful, it would be necessary to have evidence that the anxious and non-anxious groups did not show significantly different patterns of judgmental sensitivity. This was investigated by transforming the data, which had been gathered in three categories of judgment, into two categories by means of "proportional division" and by computing the reciprocal of the *SD* of the resulting distribution for each subject. The variances of these reciprocals of the *SD*s were computed for each group and subjected to Bartlett's test of homogeneity of variance. The resulting chi-square was .83, with 1 *df*, which is not significant.

The lack of differences in sensitivity other than in the number of doubtful judgments supports the contention that anxiety level, as measured by Welsh's *A* Scale, profoundly affects the tendency to report doubtful under conditions of marginal discriminability.

To check on possible serial effects in either or both experimental groups, the number of doubtful judgments in the first five series of presentations was compared with the number of doubtful judgments in the last five series of presentations; this was done by means of the *t* test for each group and for both groups combined. None of the resulting *t*'s proved to be significant and it was concluded that no serial effects were operative in either group or both groups combined.

DISCUSSION

The confirmation of the hypothesis of this study allows two conclusions: first, that the *A* Scale is sufficiently sensitive to differentiate "normal" ranges of anxious behavior, where normal is defined as a *T* score of below 70 on Welsh's *A* Scale; secondly, that the trait constellation labeled anxiety manifests itself in judging behavior in a psychophysical experiment.

Furthermore, the total number of doubtful judgments in a psychophysical experiment may be a useful measure of anxiety, as measured by Welsh's *A* Scale, provided additional research can establish the temporal stability of this response tendency as well as its reliability across sensory modalities. Also, subsequent experimentation should use populations which cover the entire score range on Welsh's scale.

The findings of this study also seem to indicate the feasibility of using a methodology such as this to try and determine the relationship between the type of behavior measured psychophysically and other underlying personality variables. The contribution of the postulated personality entities towards the behavior measured might then be determined by correlational or factorial methods.

REFERENCES

CARTWRIGHT, D. Relation of decision time to the categories of response. *American Journal of Psychology*, 1941, **54**, 174–196.

CARTWRIGHT, D., & FESTINGER, L. A quantitative theory of decision. *Psychological Review*, 1943, **50**, 595–621.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

FERNBERGER, S. W. Coefficients of precision in the method of constant stimuli. *Journal of Consulting Psychology*, 1949, **62**, 591–592.

FESTINGER, L. Studies in decision: I. Decision time, relative frequency of judgment and subjective confidence as related to the physical stimulus difference. *Journal of Experimental Psychology*, 1943, **32**, 291–306. (a)

FESTINGER, L. Studies in decision: II. An empirical test of a quantitative theory of decision. *Journal of Experimental Psychology*, 1943, **32**, 411–423. (b)

GEORGE, S. S. Attitude in relation to psychophysical judgment. *American Journal of Psychology*, 1917, **28**, 1–37.

GUILFORD, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.

KELLOG, W. A. An experimental evaluation of equality judgments in psychophysics. *Archives of Psychology*, N. Y., 1930, **17**(112), 5–78.

URBAN, F. M. The psychometric constant delta. *American Journal of Psychology*, 1950, **63**, 100–107.

WELSH, G. S. Factor dimensions A and R. In G. S. Welsh and W. G. Dahlstrom (Eds.), *Basic readings in the MMPI*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received March 12, 1965)

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 465–467

RELATIONSHIP OF ANXIETY (DRIVE) AND RESPONSE COMPETITION IN PROBLEM SOLVING[1]

JOSEPH J. TECCE[2]

National Institute of Mental Health

High (HA), medium (MA), and low (LA) anxiety groups (Taylor MA Scale) were given Duncker's candle-box problem in either the boxes-full (BF) or boxes-empty (BE) condition. Relative to the BE (low-response competition) task, the BF (high-response competition) task impaired problem-solving performance in the HA ($p < .001$) and LA ($p < .01$) groups but not in the MA group. Although these results support, in part, Hull-Spence behavior theory, the inverted U hypothesis of Yerkes and Dodson provides the more encompassing theoretical interpretation of the present findings.

Duncker's (1945) concept of functional fixity applies to a problem-solving task where a well-established response prevents a second, weaker response from occurring. For example, in the candle-box problem, Duncker showed that since the usual function of boxes is that of a container, filled boxes more strongly than empty boxes prevent the use of shelf or platform properties of boxes when the use of those properties is necessary for correct problem solving. If the "boxes-full" situation is regarded as one of high-response competition, that is, the correct-response tendency "box as a shelf" is considerably weaker than the incorrect-response tendency "box as a container" and if the "boxes-empty" situation is considered to reduce the dominance of the incorrect-response tendency "box as a container" (low-response competition), then according to Hull-Spence behavior theory (Spence, 1956), high anxiety (drive) subjects would be expected to show greater impairment in problem-solving performance for the boxes-full (high-response competition) condition as compared to the boxes-empty (low-response competition) condition. Low and medium anxiety subjects, on the other hand, should show relatively less impairment than the high-anxiety (HA) group on the boxes-full task as compared to the boxes-empty situation. There is empirical evidence for this drive-response competition interaction in a situation where experimentally produced high drive led to inferior performance in the boxes-full situation and superior performance in the boxes-empty arrangement (Glucksberg, 1962). In this latter study, however, drive was defined by monetary incentive and no medium drive condition was included. In the present experiment, drive is defined by high, medium, and low scores on the Taylor Manifest Anxiety scale (*MA* scale) in order to test the drive-response competition interaction for three levels of drive.

METHOD

Subjects. The Taylor *MA* scale was given to 190 normal volunteers and personnel at the National Institutes of Health. Sixty-eight were males; 122, females. Mean age was 24.18 ($SD = 6.92$). All subjects had some college background, the average number of years, 2. The *MA* scale scores of 18-37 ($n = 69$) constituted the high-anxiety group (HA); scores of 10-17 ($n = 71$), the medium-anxiety group (MA); and scores of 2-9 ($n = 50$), the low-anxiety group (LA). Either a K score of over 22 or an L score of over 6 resulted in exclusion of the subject.

Materials. A modification of Duncker's materials for the candle-box problem was used. Stimulus materials were three small (paper-clip size) boxes, five birthdaytype candles, five wooden matches, five thumbtacks, and a cardboard placard (simulated wall) 68 centimeters long and 36 centimeters high.

Procedure. Assignment of subjects to either the boxes-full (BF) or boxes-empty (BE) condition was determined by the toss of a coin. For the BF condition, five thumbtacks were in one box, five matches in another, and four candles in a third box.

---

[1] This work was carried out while the author was a Postdoctoral Research Fellow (MF-13,914) at the National Institute of Mental Health. The collection of data by Mary Houser and Louise Scott, the statistical aid of Velma Harris and Ruth McClelland, and the general management of administrative problems by Audrey Jordan are gratefully acknowledged.

[2] Now at the Division of Psychiatry, Boston University Medical Center.

TABLE 1

NUMBER OF SUBJECTS IN EACH ANXIETY GROUP PASSING AND FAILING IN BF AND BE CONDITIONS

Anxiety group	Condition of boxes	Number of subjects		χ^2
		Passed	Failed	
High	Full	17	14	10.84**
	Empty	35	3	
Medium	Full	27	11	1.22
	Empty	28	5	
Low	Full	16	12	7.51*
	Empty	21	1	

* $p < .01$.
** $p < .001$.

The fifth candle lay outside the box as the one to be mounted. For the BE situation, thumbtacks, matches, and candles lay outside their respective boxes. All subjects were asked to verbalize solution of the problem of mounting vertically a single candle on the cardboard "wall" using any of the materials in front of them. Following these instructions, stimulus materials were displayed by removal of a large cardboard cover. The acceptable solution involved tacking one of the three boxes to the wall and using it as a shelf or platform on which to place a lighted candle. Verbalization of the problem's solution or failure to give the correct solution in 15 minutes terminated the experiment. Following the candle problem, the subject completed the *MA* scale.

RESULTS

The first finding concerns the relative difficulty of experimental conditions. A smaller proportion of subjects reached solution in the BF condition than in the BE condition. For BF, 60 subjects reached solution or passed (P) versus 37 failed (F); for BE, there were 84 P versus 9 F. A two-sample χ^2 of 19.45 for these data was significant at the .001 level, indicating a greater proportion of solutions in the BE condition. Thus, the BF task was more difficult than the BE task.

A second group of findings involves the relative performance of anxiety groups in the two experimental conditions. As can be seen in Table 1, a significantly smaller proportion of LA and HA subjects reached solution in the BF situation as compared with the BE situation. For the HA group, there were 17 P and 14 F in the BF condition versus 35 P and 3 F in the BE situation ($\chi^2 = 10.84$, $p < .001$). Similarly, for the LA group, there were 16 P and 12 F in the BF condition versus 21 P and 1 F in the BE arrangement ($\chi^2 = 7.51$, $p < .01$). Therefore, HA and LA groups performed significantly worse in the BF condition than in the BE condition. On the other hand, there was no difference in the proportions of MA subjects reaching solution in the BF (27 P and 11 F) and BE (28 P and 5 F) situations ($\chi^2 = 1.22$, $p > .20$). Thus, of the three anxiety groups, only the MA group was comparable in performance on the BE and BF tasks. Therefore, relative to the BE or low-response competition task,[3] the BF or high-response competition situation impaired problem-solving performance in the HA and LA groups but not in the MA group.

DISCUSSION

The finding that relatively fewer subjects in the HA group reached solution in the high-response competition (BF) as compared to the low-response competition (BE) condition supports Hull-Spence behavior theory and is in agreement with the previous finding of greater functional fixity in high-drive subjects (Glucksberg, 1962). A similar finding for the LA group, however, that is, a smaller proportion of subjects reaching solution in the BF than in the BE condition, was not expected and suggests that either the Taylor *MA* scale is an imprecise measure of drive or that problem solving, specifically the candle-box problem, does not fall within the boundary conditions of Hull-Spence theory. The latter possibility is an unlikely one, however, since the Glucksberg study supported Hull-Spence theory using experimentally induced high and low drive. Therefore, the lower end of the *MA* scale does not seem to measure drive alone. Since LA subjects usually have high *K* scores (*A* and *K* scores on the *MA* scale are negatively related), high defensiveness appears to be related to impairment in problem solving.

Of the three anxiety groups, only the MA subjects did not show an impairing effect of high-response competition relative to low-response competition. This finding suggests that a moderate level of drive is optimal for difficult problem solving and is in agreement with previous work (Tecce & Testa, 1965) supporting the inverted U hypothesis (Yerkes & Dodson, 1908) that postulates a nonmonotonic relation between arousal (or drive) and task difficulty. According to this view, problem-solving inferiority of the LA and HA groups was caused by under- and over-arousal, respectively, that resulted from anxiety level and high task difficulty in the BF condition. The MA group, on the other hand, was optimally aroused in the BF task. In conclusion, the most encompassing theoretical interpretation of the present findings is the Yerkes-Dodson hypothesis.

---

[3] In the BE condition, there is no difference in the proportion of solutions among the HA (35 P; 3 F), MA (28 P; 5 F), and LA (21 P; 1 F) groups ($\chi^2 = 2.45$, $p > .20$). Therefore, performance measures in the BE condition seem to be appropriate control data against which BF performance was compared in the chi-squares presented in Table 1.

REFERENCES

DUNCKER, K. On problem-solving. *Psychological Monographs,* 1945, 58(5, Whole No. 270).

GLUCKSBERG, S. The influence of strength of drive on functional fixedness and perceptual recognition. *Journal of Experimental Psychology,* 1962, 63, 36–41.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

TECCE, J. J., & TESTA, DONNA H. Effects of heightened drive (shock) on 2-, 4-, and 8-choice card-sorting. *Perceptual & Motor Skills,* 1965, 20, 715–716.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative and Neurological Psychology,* 1908, 18, 459–482.

(Received March 15, 1965)

ERRATUM

On page 315 of the article "Commonality and Stability of Word Association Responses in Good and Poor Premorbid Schizophrenics," by Paul R. Dokecki, Lewis G. Polidoro, and Rue L. Cromwell (*Journal of Abnormal Psychology,* 1965, 70, 312–316), there are statistical inaccuracies in the section entitled *Stability Scores* (*SS*). In the simple randomized design analysis of variance the F and t values should have been larger. Specifically, $F = 4.10$ ($df = 2/51$, $p < .025$). The inequalities remain the same: TB = GPS; TB > PPS ($t = 2.47$, $df = 51$, $p < .02$, one-tailed); GPS > PPS ($t = 2.47$, $df = 51$, $p < .02$, one-tailed). In the factorial design analysis of variance, the interaction should not have been reported as significant ($F < 1$) and simple effects should not have been reported. Instead the main effect of groups, as in the simple randomized design, was significant, as was the response style effect, showing C > I on stability ($F = 9.51$, $df = 1/48$, $p < .005$). It should be pointed out that these corrections in no way alter the overall results and their interpretation (in the originally reported interaction there was no deviant cell). The last two sentences in the abstract should be ignored in favor of the main effects reported above.

Journal of Abnormal Psychology
1965, Vol. 70, No. 6, 468–472

INTERPERSONAL TYPES AMONG PSYCHIATRIC PATIENTS

MAURICE LORR, PATRICIA F. BISHOP

Veterans Administration, Washington, D. C.

AND DOUGLAS M. McNAIR

Boston University

The aims of the study were to isolate some of the main interpersonal behavior types within a sample of nonpsychotic psychiatric patients in psychotherapy, to determine the constancy of the isolated types across the 3 subsamples of 150 cases, and to identify any systematic differences among them. Each patient was described by his therapist on the 15 factors of the Interpersonal Behavior Inventory. The 4 replicated types were (a) inhibited, submissive, abasive; (b) agreeable, nurturant, sociable; (c) hostile, mistrustful, detached; (d) dominant, competitive, exhibitionistic. Some diagnostic and occupational differences were found.

Recent literature has emphasized the role of faulty interpersonal behaviors in the psychoneurotic and personality disorders. Indeed to Horney (1945), Fromm (1947), and Sullivan (1947), the milder behavior disorders represent mainly problems in relating to people. A similar view has been taken by Leary (1957) and his colleagues Coffey, Freedman, Ossori, and LaForge. They have sought to develop a scheme for multilevel interpersonal diagnosis. Conventional psychiatric diagnoses, on the other hand, are described mainly in terms of patient symptoms and complaints. The psychoneurotic disorders are defined principally by symptoms in *Mental Disorders* (APA, 1952). However, interpersonal characteristics receive somewhat greater emphasis in definition of the personality disorders.

The view taken here is that diagnoses of the milder disorders should be based primarily on characteristic interpersonal behavior patterns. Associated neurotic symptoms and complaints should be regarded as secondary to maladaptive patterns of interpersonal adjustment. The present study was undertaken to explore the use of currently observable interpersonal behavior as the primary criterion of classification. Specifically the study aims were (*a*) to isolate the major classes of homogeneous interpersonal profiles within a sample of nonpsychotic psychiatric patients, (*b*) to determine the constancy of the classes isolated across three subsamples, and (*c*) to identify any systematic differences between the classes isolated with respect to background characteristics.

METHOD

Rating Instrument

The basic measuring device was the Interpersonal Behavior Inventory (IBI-3). The development of the IBI, its standardization, factor structure and reliability, and a theory of the organization of its 15 categories are presented elsewhere (Lorr & McNair, 1963, 1965). The inventory consists of 160 statements descriptive of manifest interpersonal behaviors. The format permits the rater to indicate how often the person rated exhibits the behavior in question: not at all, occasionally, fairly often, or quite often. A category score is the sum of 7 to 11 unit-weighted items defining each factor. The interpersonal categories are labeled as follows: Dominance, Recognition, Hostility, Mistrust, Detachment, Inhibition, Abasiveness, Submissiveness, Succorance, Deference, Agreeableness, Nurturance, Affection, Sociability, and Exhibition. Table 1 presents illustrative examples of each of the categories.

Sample

The patient sample consisted of 212 men and 313 women being seen in individual psychotherapy at least once a week for a minimum of 3 months. The therapists were instructed to rate a sample of their case load as diverse as possible with respect to interpersonal behavior, occupation, and social class. All rated were to be nonpsychotic. Actually the typical patient had received 95 therapy interviews. Approximately 92% were high-school graduates and 68% had received some college training. The average patient was 33 years old; 72% were married. Approximately 48% were diagnosed psychoneurotic, 37% were labeled personality disorders, and 16% were categorized otherwise.

The therapists' sample consisted of 104 psychologists and 12 psychiatrists. Most were members of the Clinical Division of the American Psychological Association and a majority had been certified by ABEPP. The cases each reported were being seen in private practice, and in university, community, and state outpatient clinics. Therapists were paid a modest fee for each patient rated. While no claim is made for representativeness, the therapists included did come from all parts of the country and were uniformly well qualified.

Typing Process

To establish a common scale, the raw factor scores were converted into standard scores with a mean of five and a standard deviation of one. Thus each individual was represented by 15 derived standard scores. The index of similarity used was the correlation coefficient. A correlation coefficient reflects primarily similarity in the shapes of two profiles. However, in the process to be described, it is always possible to bring in level as well, after profiles alike in shape have been grouped. In contrast, a distance measure (*D*), advocated by some, has no unique meaning. It may reflect a large difference between two profiles on one score, or dimension, or the sum of small differences on all dimensions. The *D* measure also fails to take the direction of differences into account; it fails to differentiate between profiles with opposite slopes.

The typing process begins by listing with each coded profile, the code numbers of all other profiles that correlate with it .55 or higher. At this value a coefficient is significant at the 5% level and two profiles are defined as "similar." The profile with the longest associated list is selected as a pivot. To the pivot is added the profile with the highest average correlation with all members in the pivot list. To the first pair is added the profile that correlates highest on the average with both. Other profiles are added to the group of profiles by the same procedure until none correlates on the average above .55 with other group members. Next, any profile in the residual matrix that correlates above .40 (on the average) with the first group or type is deleted. The second type is then evolved from the residual matrix by first listing with each profile all other profiles that correlate .55 or better with it. The pivot and the members of the second cluster are selected in the manner previously described. Again profiles correlating above .40 with the second group are deleted from the matrix. All later types are formed by the same selection procedure described. The final step involves computing the mean correlation between each member of a given type with all members of all the other types, each type being taken separately. If the profile correlates (on the average) above .40 with another type it is removed. Any type with less than four members is dropped.

Each type is then characterized by its mean profile and by the highest and lowest score on each of the categories defining it. Members of a type are those individuals whose standard scores lie within each and every bound set by the highest and lowest scores. Thus defined the classes of homogeneous profiles or types are mutually exclusive.

TABLE 1

STATEMENTS EXEMPLIFYING THE FIFTEEN INTERPERSONAL CATEGORIES OF IBI-3

Category	Statements
Dominance	Bosses his friends and associates around. Takes charge of things when he's with people.
Recognition	Seizes opportunities to rival and surpass others. Strives for symbols of status and superiority to others.
Hostility	Ridicules, belittles or depreciates others. Uses a sarcastic or biting type of humor.
Mistrust	Mistrusts the intentions of others toward him. Expresses suspicion when someone is especially nice to him.
Detachment	Acts business-like and impersonal with co-workers. Keeps aloof from his neighbors.
Inhibition	Shows discomfort and nervousness when people watch him at work or at play. Shows signs of self-consciousness with strangers.
Abasiveness	Blames himself when interpersonal friction with others occurs. Apologizes for not having done better when he completes a task.
Submissiveness	Gives in rather than fight for his rights in a conflict. Lets his friends or spouse push him around.
Succorance	Asks others to look after his interests. Asks for help on jobs he could handle himself.
Deference	Carries out orders of his superiors with zest. Takes the role of helper or supporter of authority figures.
Agreeableness	Contributes positively as a member of some group or team. Relates to and treats people as equals.
Nurturance	Listens sympathetically to others talk about their troubles. Puts aside his own work or pleasure if someone asks for help.
Affection	Shows a real liking and affection for people. Acts close and personal with people.
Sociability	Invites friends and acquaintances to his home. Drops in to visit friends just to socialize.
Exhibition	Draws attention to himself in a group by telling jokes and anecdotes. Acts the clown or amuses others at a party.

TABLE 2

AVERAGE CORRELATIONS WITHIN AND BETWEEN CLUSTERS FOR SETS 1, 2, AND 3

Relation	Set 1	Set 2	Set 3
1 versus 1	65	68	71
1 versus 2	−32	−52	−14
1 versus 3	01	−15	03
1 versus 4	−64	07	−64
2 versus 2	72	68	66
2 versus 3	−59	−08	−64
2 versus 4	13	−01	09
3 versus 3	65	65	66
3 versus 4	17	−59	01
4 versus 4	68	63	67

The typing process described has been programmed for data processing equipment and can handle as many as 150 variables at one time. Because only 150 cases could be processed at one time 3 subsamples of 150 cases each were selected after the entire sample had been randomized. Randomnization was necessary because the inventories received did not come from all sources at equal rates.

RESULTS

The analysis yielded four patient types in two of the subsamples and five in the third. Because the fifth type in the last subsample consisted of only four members and remained unconfirmed, it will not be discussed. The average correlations within and between types are given in Table 2. It is clear that the positive between-group correlations are all low, none being higher than .13. The within-group correlations are relatively high indicating adequate similarity between members.

Table 3 presents evidence of the degree of constancy in patient type from one subsample to the next. The coefficient of congruency (Harman, 1960) was employed as a measure of the degree of similarity between the mean profiles of the four types. The subsample types are coded by subsample and type. For example, type 1.1 is the first type found in the first subsample. If a coefficient of .75 indicates adequate similarity and a coefficient of .90 marked similarity then all of the types have clear matches in each of the subsamples.

Each type will be described in terms of those elements of its profile that average at least one standard deviation above the overall mean for the entire patient sample. By this criterion (see Table 4), Type I is primarily an inhibited, submissive, and abasive group. The type may be further characterized by the absence of Dominance, Competitiveness, and Hostility. As will be seen shortly this type is most likely to be labeled neurotic and to be accepted for psychotherapy. Type II is characterized by Agreeableness, Nurturance, Affection, and Sociability. As might be anticipated, Hostility, Mistrust, and tendencies toward Detachment are low. Leary (1957) describes a similar patient group whom he characterizes as sweet, bland, conventional, and hysteric.

Type III is defined primarily by high Hostility, Mistrust, and Detachment from others. Agreeableness, Nurturance, Love, and Socia-

TABLE 3

CONGRUENCIES BETWEEN MEAN CLUSTER PROFILES

Type	Subsample	Type											
		I			II			III			IV		
		1.1	2.4	3.2	1.2	2.1	3.4	1.3	2.2	3.1	1.4	2.3	3.3
I	1.1												
	2.4	90											
	3.2	94	88										
II	1.2	−41	−46	−45									
	2.1	11	01	13	78								
	3.4	18	13	19	77	97							
III	1.3	03	04	05	−93	−86	−92						
	2.2	01	01	13	−80	−72	−88	90					
	3.1	−14	−25	−17	−68	−80	−90	91	92				
IV	1.4	−91	−82	−94	20	−39	−44	21	09	39			
	2.3	−94	−89	−98	48	−10	−15	−10	−17	15	92		
	3.3	−92	−86	−93	57	−02	−05	−21	−28	06	87	97	

Note.—Decimals have been omitted.

bility are rated low. It seems likely that Type III would ordinarily be designated Schizoid. The last replicated type (IV) is predominantly exhibitionistic, dominant, competitive, and hostile. Naturally such patients score low on Inhibition, Submissiveness, and Abasiveness.

Type Differences

The types were compared with respect to age, marital status, and highest grade completed. The only significant difference between the types was with respect to age; Type II members were older as a group. The mean age of the types ranged from 32 to 36. The percentage married ranged from 59 to 79. The percentage with some college education ranged from 62 to 74. Thus background variables appeared not to relate closely to the patient types.

Table 5 presents the therapist's diagnostic impressions grouped by major disorder. A chi-square test indicated differences among the cells significant at $p < .01$. Type I members (self-effacing and submissive) compared to other types tend to be diagnosed as psychoneurotic or personality pattern disorders. Diagnosis appears not to differentiate Type I from Type III (withdrawn, hostile, suspicious). Type II members (agreeable, helpful, responsible, friendly) tend to be labeled psychoneurotic or to be left undiagnosed. Personality pattern diagnoses (Schizoid or Paranoid Personality) are less frequent for Type II than for the others. Type IV members (dominant, competitive, exhibitionistic) compared to other classes, are more frequently diagnosed personality trait disturbances, that is, Passive-Aggressive, Aggressive, or Compulsive.

TABLE 4

MEAN STANDARD SCORES ON FIFTEEN INTERPERSONAL CATEGORIES OF IBI-3

Category	Type I	Type II	Type III	Type IV
Dominance	3.89	5.13	4.91	6.61
Recognition	3.97	4.84	5.19	6.34
Hostility	3.91	4.12	6.06	6.13
Mistrust	4.49	3.88	6.08	5.51
Detachment	5.68	4.09	6.10	4.14
Inhibition	5.98	4.10	5.69	3.93
Submissiveness	6.17	4.79	4.44	3.91
Succorance	5.25	4.27	4.96	4.85
Abasement	6.02	4.61	4.44	4.02
Deference	5.31	5.62	4.11	4.74
Agreeableness	5.99	6.44	3.82	4.66
Nurturance	5.11	6.20	3.92	4.83
Affection	4.87	6.24	3.88	5.09
Sociability	4.34	5.85	3.97	6.03
Exhibition	4.15	4.69	4.75	6.59

TABLE 5

TYPE MEMBER DISTRIBUTION BY OCCUPATION IN PERCENTAGES

Occupation	Type I	Type II	Type III	Type IV
Administration	2	9	8	13
Teaching	8	12	12	5
Professional	12	17	10	21
White collar	18	23	21	8
Manual worker	6	7	14	8
Housewife	43	21	25	33
Student or unemployed	10	12	8	13

TABLE 6

PATIENT DIAGNOSIS BY TYPE

Disorder	Type I	Type II	Type III	Type IV
Psychoneurotic	26	25	22	10
Personality pattern	11	2	11	5
Personality trait	9	6	12	19
Other	3	10	3	5
Total	49	43	48	39

Further insight into the character of the types can be gained from their occupations. Type I members tend to be housewives. Type II members tend to come from any of the white collar occupations and are least frequently housewives. On the other hand, Type III members (withdrawn, suspicious) tend to come from unskilled and semiskilled blue-collar occupations. Type IV members, consistent with their aggressive and competitive interpersonal profiles, tend to be in administration, management or in one of the professions.

DISCUSSION

Some comparison of present findings with those from a prior study (McNair & Lorr, 1965) seems warranted. In the initial study a sample of 450 veterans in outpatient psychotherapy were described by their therapists

($N = 259$) on an earlier rating form called IBI-2. The inventory measured 11 rather than 15 dimensions of interpersonal behavior. A type analysis essentially the same as described here was employed. However, the criterion for separation of types was less exacting. Of the eight patient types isolated in the earlier analysis, three appear to have a good match among the types reported here. The McNair-Lorr Type B (dominant, hostile, mistrustful) appears to be very similar to Type IV. A Type C (nurturant, affiliative, sociable) resembles Type II, while Type D (hostile, detached, mistrustful) seems a close match to Type III. Finally a Type F (passive, abasive, nurturant) partially resembles Type I. The evidence, although qualitative in nature, tends to lend support to the viability of the types reported here.

An important applied question relates to decision rules for assigning cases to one of the four types with the fewest number of misclassifications. A solution to this problem was not pursued because of the exploratory nature of the study. The clustering process classified between 48% and 49% of the 150 members of each sample to one of the types. These results suggest that rules could be established to assign a high proportion of all cases without excessive error.

The interpersonal behavior factors of nonpsychotic but disturbed individuals have been shown to be essentially no different from those of normals (Lorr & McNair, 1965). Interpersonal tendencies in the individual needing psychiatric help are principally more extreme or accentuated. Why then is the characteristic interpersonal behavior profile more important diagnostically than the symptom grouping? One answer is that interpersonal variables are more crucial and relevant therapeutically. Therapists find such variables are more predictive of expected patient behavior in group or individual treatment. Also the ability of the individual to form a relationship and the nature of the relationship may be anticipated better with such knowledge (Leary & Coffey, 1955).

Characteristic interpersonal "styles" or ways of relating to peers, authority figures and to subordinates are relatively enduring characteristics. Interpersonal style in turn influences, and is influenced by, preferred defense mechanisms for resolving inner conflict. Accordingly it is plausible to hypothesize that the submissive-abasive-inhibited type tends to turn against himself, while the detached-hostile-suspicious type turns against others and withdraws. Perhaps the friendly-agreeable-helpful type tends to use denial and repression. In the dominant-competitive-exhibitionistic individual we may find use of acting out or displacement.

In conclusion, it can be said that the method of typing yields meaningful, mutually exclusive and replicable outpatient types. Relationships between the types and conventional psychiatric diagnosis are not very strong. However, as the criteria of classification are quite different in the two systems of classification, close correspondence should not be expected. On the whole, when one considers the exploratory nature of the study, the approach and the findings show promise for a useful and fruitful system of patient classification.

REFERENCES

AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and statistical manual for mental disorders.* Washington, D. C.: Mental Hospital Service, 1952.

FROMM, E. *Escape from freedom.* New York: Rinehart, 1947.

HARMAN, H. H. *Modern factor analysis.* Chicago: Univer. Chicago Press, 1960.

HORNEY, KAREN. *Our inner conflicts.* New York: Norton, 1945.

LEARY, T. *Interpersonal diagnosis of personality.* New York: Ronald Press, 1957.

LEARY, T., & COFFEY, H. S. Interpersonal diagnosis: Some problems of methodology and validation. *Journal of Abnormal and Social Psychology,* 1955, **50**, 110-124.

LORR, M., & MCNAIR, D. M. An interpersonal behavior circle. *Journal of Abnormal and Social Psychology,* 1963, **67**, 68-75.

LORR, M., & MCNAIR, D. M. Expansion of the interpersonal behavior circle. *Journal of Personality and Social Psychology,* 1965, 2.

MCNAIR, D. M., & LORR, M. Differential typing of psychiatric outpatients. *The Psychological Record,* 1965, **15**, 33-41.

SULLIVAN, H. S. *Conceptions of modern psychiatry.* Washington, D. C.: William Alason White Psychiatric Foundation, 1947.

(Received March 17, 1965)